Two Sample Unconditional Quantile Effect

Atsushi Inoue [*]

Vanderbilt University

Tong Li [†]

Vanderbilt University

Qi Xu [‡]

Vanderbilt University

Abstract

This paper proposes a new framework to evaluate *unconditional quantile effects* (UQE) in a data combination model. The UQE measures the effect of a marginal counterfactual change in the unconditional distribution of a covariate on quantiles of the unconditional distribution of a target outcome. Under rank similarity and conditional independence assumptions, we provide a set of identification results for UQEs when the target covariate is continuously distributed and when it is discrete, respectively. Based on these identification results, we propose semiparametric estimators and establish their large sample properties. Applying our method to a variant of Mincer's earnings function, we study the counterfactual quantile effect of actual work experience on income.

Keywords: Counterfactual policy effect, unconditional quantile effect, data combination model, Mincer regression.

JEL Classification: C10, C31.

---

[*]Department of Economics, Vanderbilt University. Email: atsushi.inoue@vanderbilt.edu

[†]Department of Economics, Vanderbilt University. Email: tong.li@vanderbilt.edu

[‡]Department of Economics, Vanderbilt University. Email: qi.xu.1@vanderbilt.edu

1 Introduction

Missing data is a ubiquitous problem in empirical studies. Consider the scenario where a researcher is interested in conducting counterfactual analysis on a target variable, but it is entirely missing from the dataset of interest. In such circumstances, counterfactual policy effects cannot be identified from the primary dataset alone, and therefore, external information and/or stronger identifying assumptions are necessary. In this paper, we utilize both to achieve identification. Specifically, we focus on the situation where the missing variable can be found in another dataset and the information from which can be used to recover target policy parameters in the population of interest, under a set of commonly assumed restrictions on both the data structure and the model primitives.

To fix ideas, consider the following example. Suppose we are interested in studying the effect of a counterfactual change in the distribution of actual labor market experience on some distributional features of yearly earnings. Our main dataset does not record respondents' work history, and therefore, we cannot recover their actual labor market experience. Suppose the variable is available from a second dataset, but it may not be a reliable source of information on income or it may not be representative of the target population we aim to analyze. In this case, we would benefit from combining information from both samples to identify and estimate our parameter of interest.

Research on counterfactual policy effects under data combination is scarce. Our paper fills this gap by proposing a new framework that accommodates such a data structure. In this paper, we focus on one particular type of counterfactual policy effects, the *unconditional quantile effect* (UQE). It measures the effect of a marginal change in the unconditional distribution of a single covariate on the quantiles of a target outcome. We provide identification results for UQE under various types of marginal distributional change. The key insight of our identification strategy is that some covariates present in both datasets can be excluded from the outcome equation, which would provide a source of exogenous variations that allows us to recover the joint distribution of missing variables, otherwise not identified using the two samples separately.

The second contribution of the paper is to propose novel semiparametric estimators based on these identification results. Departing from the literature on the estimation of counterfactual quantile effects—see, e.g. Firpo et al. (2009b), Sasaki et al. (2022), etc.—which focuses primarily on the *marginal location shift* (MLS) of a covariate, we provide estimators of UQE under two

general types of counterfactual distributional changes, namely the *marginal distributional shift* (MDS) and the *marginal quantile shift* (MQS),[1] the latter of which includes MLS as a special case. To the best of our knowledge, large sample results for these two cases are new to the literature. We apply these results to study a variant of Mincer's earnings function. Using data from Integrated Public Use Microdata Sample (IPUMS) as our main data source and the Panel Study of Income Dynamics (PSID) as the auxiliary sample, we investigate the counterfactual effect of actual work experiences on income. The effect profiles with MDS and MQS are found to be similar in shape.

This paper belongs to the growing literature on the (marginal) unconditional policy effect. Since Firpo et al. (2009b) introduced the method of *unconditional quantile regressions* (UQR), the study of unconditional policy effect has gained much attention. In general, this parameter differs from the one identified by the conditional quantile regression (Koenker and Bassett Jr, 1978), where marginal effects on the conditional quantile are the locus of attention. Applied researcher are often interested in the shifts in the quantiles of unconditional distribution of a target outcome. For instance, one may take an interest in how wage distribution changes in response to marginal increases in some characteristics of the labor force, such as education level and experience. Conditional quantile regression cannot be applied to address this type of questions, whereas UQR suits the goal.

Rothe (2012) generalizes the method of Firpo et al. (2009b), and analyzes a variety of counterfactual policy effects. He formalizes the idea of *ceteris paribus* distributional change and provides extensive results for both fixed and marginal policy shifts. Our identification framework is closely related to his treatment of the latter type. Focusing on the special case of quantile effects, we extend his identification results to a data combination setting and provide novel inference theories specifically tailored to the distinct features of combined samples. For recent development in this literature, see Firpo et al. (2018), Martinez-Iriarte and Sun (2020), Martínez-Iriarte (2021), and Sasaki et al. (2022). For a comprehensive survey on counterfactual distributions and decomposition methods, see Fortin et al. (2011).

Our paper also builds on the econometric methods of data combination. In economics, this strand of literature stems from the *two-sample instrumental variables* (TSIV) model that was first introduced by Klevmarken (1982), Angrist and Krueger (1992), Arellano and Meghir (1992), and is later extended by Ridder and Moffitt (2007), Inoue and Solon (2010), among

---

1 The precise definitions of MLS, MDS, and MQS are given in Section 3.

others. Conceptually, the semiparametric data combination model we consider here is different from the traditional missing data problem (Robins et al., 1994). It is more closely related to the "verify-out-of-sample" model in Chen et al. (2008), and also to Imbens and Lancaster (1994), Fan et al. (2014), Graham et al. (2016), Hirukawa et al. (2020), and Buchinsky et al. (2022), to name a few.

The paper is organized as follows. In the next section, we describe the model and assumptions on the data structure. In Section 3 we introduce the parameter of interest, and then present identification results for continuously distributed and discrete target covariates, respectively. Section 4 discusses the estimation strategy and large sample results. We apply the method to study the income effect of real labor market experience in Section 5. Section 6 concludes.

2 Setup

The objective of our paper is to analyze the effect of a counterfactual change in the marginal distribution of the covariate of interest, X, on the quantiles of the target outcome, Y, under data combination. The precise definition of the counterfactual policy effect is provided in Section 3. When X is exogenous, and all the variables relevant for analysis are observed from a single data source, counterfactual policy effects can be analyzed either directly by applying tools from Firpo et al. (2009b) and Rothe (2012), or indirectly by recovering the structural function using standard identification results such as Matzkin (2003) and Matzkin (2007). However, when the variables of interest are scattered among several different data sources, we face a fundamental identification problem: The conditional distribution of Y given X is not identified from any single sample. In this case, existing methods do not provide an immediate solution.

Throughout this paper, we consider the scenario where our Y and X are sourced from two different data sets. The outcome is contained in the *principal* or *main* sample, $\mathcal{S}_s = \{Y_i, Z_i\}_{i=1}^{n_s}$, from the *study* population, $\mathcal{P}_s$. The target covariate is missing completely from $\mathcal{S}_s$. However, it is observed in the *auxiliary* sample, $\mathcal{S}_a = \{X_i, Z_i\}_{i=1}^{n_a}$, from the *auxiliary* population, $\mathcal{P}_a$, which does not contain observations of Y.

We now formally describe our structural model. We allow variables from two populations to be determined by different mechanisms. For the study population,

$$Y_s = g_s(X_s, Z_1, \epsilon_s), \tag{1}$$

4

$$X_s = h_s(Z, \eta_s), \tag{2}$$

where $Y_s \in \mathcal{Y} \subset \mathbb{R}$ is the potential outcome in the study population, $\epsilon_s \in \mathcal{E} \subset \mathbb{R}^{d_\epsilon}$ is a vector of unobserved heterogeneity term. Equation (1), links the target outcome, a scalar variable, $X_s \in \mathcal{X} \subset \mathbb{R}$, and a vector of exogenous variables, Z_1. Here, X_s is the potential covariate of interest in the study population, which is in turn determined by (2). We can think of (2) as the reduced form relationship between X_s and Z, where $Z' := (Z_1', Z_2')' \in \mathcal{Z} := \mathcal{Z}_1 \times \mathcal{Z}_2 \subset \mathbb{R}^{d_z}$ includes both the exogenous variables in the outcome equation and a vector of excluded instrument, Z_2. The vector of instrument, Z, is available in both samples, and therefore, it serves to establish a link between two samples. The model in (1) accommodates general nonseparability between covariates and the unobserved heterogeneity. We do not impose any parametric or shape restriction on g_s.

Variables in the auxiliary population are determined by

$$Y_a = g_a(X_a, Z_1, \epsilon_a) \qquad \text{and} \qquad X_a = h_a(Z, \eta_a),$$

where g_a and h_a are generally different from g_s and h_s, respectively.

Let R denote the sample membership indicator. That is, $R_i = 1$, if i-th draw comes from the study population, $i = 1, ..., n := n_s + n_a$. Let $Y := RY_s + (1-R)Y_a$ and $X := RX_s + (1-R)X_a$. If no variable is missing, we are able to observe (Y_s, Y_a, X_s, X_a). However, in our context, only RY and $(1-R)X$ are observed. We then construct a pseudo-merged sample $\mathcal{S}$ using the two data sources as $\mathcal{S} = \{R_i, R_iY_i, (1-R_i)X_i, Z_i\}_{i=1}^n$. Let $A := (R, RY, (1-R)X, Z)$ and $W := (X, Z_1)$ collect the observed variables and the covariates in the outcome equation, respectively. Throughout the paper, we arrange the data in such a way that $R_i = 1$ for $i = 1, ..., n_s$ and $R_i = 0$ for $i = n_s + 1, ..., n$. The merged sample may not correspond to any real-world population. We impose the following set of assumptions on the merged sample so it can mimic a random sample from a pseudo population. These assumptions are largely based on Assumption 1 in Graham et al. (2016).

Assumption 1 (Data Structure)

(a) $\text{Supp}(F_{Z|R=1}) \subset \text{Supp}(F_{Z|R=0})$.

(b) (i) $n_s/(n_s + n_a) \to Q_0$; (ii) R follows a Bernoulli distribution, with $\mathbb{E}[R] = Q_0$.

(c) There is a unique measurable function $r(\cdot) : \mathcal{Z} \mapsto [0,1]$, such that for all $z \in \mathcal{Z}$,

$$\frac{f_{Z|R}(z|1)}{f_{Z|R}(z|0)} = \frac{1-Q_0}{Q_0}\frac{r(z)}{1-r(z)}.$$

(d) (i) $Q_0 \in (\epsilon_1, 1-\epsilon_1)$, for some $\epsilon_1 \in (0,1/2)$; (ii) $\epsilon_2 < r(z) < 1-\epsilon_2$ for some $\epsilon_2 \in (0,1/2)$, and for all $z \in \mathcal{Z}$.

(e) $(X_s|Z, R=1) \overset{d}{=} (X_a|Z, R=0)$.

Assumption 1(a) is a support condition on the commonly observed variables. It ensures that we will be able to find, for all the observations in the study sample, comparable units in the auxiliary sample, Assumption 1(b) imposes a pseudo randomization scheme on R, and therefore, allows us to view the merged data as a random sample from the pseudo-merged population. Let $\ell(\cdot)$ denote the conditional likelihood ratio of Z across two population, i.e. $\ell(z) := f_{Z|R}(z|1)/f_{Z|R}(z|0)$. Assumption 1(c) expresses this likelihood ratio as a function of $r(\cdot)$, which plays the role of the "propensity score" function of R given Z. In our context, this is the probability that one observation belongs to the study population conditional on the value that instrumental variables take. The first part of Assumption 1(d) indicates that n_s grows at the same order of magnitude as n_a. The second part of Assumption 1(d) ensures that the pseudo-true merged population is not a degenerate one conditional on all possible values of Z. By Assumption 1(b)–(d) and Bayes' Law, we have $r(z) = \mathbb{P}(D=1|Z=z)$, and thus, $r(\cdot)$ can be viewed as the propensity score function.

Assumption 1(e) is a rank similarity condition. It requires the conditional distribution of X_s given Z in the principal population coincide with that of X_a in the auxiliary population. In view of the structural relation in (2), the assumption is satisfied if $h_s = h_a$ and $(\eta_s|Z, R = 1) \overset{d}{=} (\eta_a|Z, R=0)$. Assumption 1(e) is the only cross-population restriction we impose on our data structure, which means the conditional distribution of Y given (X, Z), and therefore, the conditional distribution of Y given Z and the marginal distribution of Z are all allowed to differ across $\mathcal{P}_s$ and $\mathcal{P}_a$. This assumption is weaker than Assumption 1(ii) of Graham et al. (2016), as we do not impose a rank similarity condition on the outcome, which would imply $F_{Y_s|ZR=1}(\cdot|\cdot) = F_{Y_a|ZR=0}(\cdot|\cdot)$.

3 Identification

In this section, we first introduce the definition of UQE. Then, we develop a set of identification results, for the cases when X is continuously distributed, and when it is discrete, respectively.

3.1 Parameter of Interest

Our definition of the unconditional policy effect depends on the notion of a counterfactual experiment, which is formally defined as follows,

Definition 1 (Counterfactual Experiment) *Let $\phi := (\widetilde{\mathcal{U}}_s, \widetilde{G}_s, \widetilde{Z}, \widetilde{R}, \widetilde{\epsilon}_s, \widetilde{g}_s) : \Omega \mapsto K([0, 1]) \times \mathcal{D}(\mathcal{X}) \times \mathcal{Z} \times \{0, 1\} \times \mathcal{E} \times l_2(\mathcal{X}, \mathcal{Z}_1, \mathcal{E})$, where $K([0, 1])$ is the collection of all non-empty closed subsets of the unit interval, and $\mathcal{D}(\mathcal{X})$ denotes the space of distribution functions on $\mathcal{X}$. We say Φ is the set of counterfactual experiments, if for all $\phi \in \Phi$, we have (i) $\widetilde{G}_s^{-1}(U_s) = \widetilde{G}_s^{-1}(U_s')$ almost surely for all $U_s, U_s' \in \widetilde{\mathcal{U}}_s$; (ii) $(\widetilde{\epsilon}_s, \widetilde{Z}, \widetilde{R}) \overset{d}{=} (\epsilon_s, Z, R)$; (iii) $\widetilde{g}_s = g_s$, (iv) for all $U_s \in \mathcal{U}_s$ and $\widetilde{U}_s \in \widetilde{\mathcal{U}}_s$, there exists $\widetilde{U}_s' \in \widetilde{\mathcal{U}}_s$ and $U_s' \in \mathcal{U}_s$, respectively, such that $(\widetilde{U}_s'|\widetilde{Z}_1, \widetilde{R} = 1) \overset{d}{=} (U_s|Z_1, R = 1)$ and $(\widetilde{U}_s|\widetilde{Z}_1, \widetilde{R} = 1) \overset{d}{=} (U_s'|Z_1, R = 1)$, where $\mathcal{U}_s = \{\breve{U} \in \mathcal{U}[0, 1] : (F_{X_s|R}^{-1}(\breve{U}_s|1)|Z_1, R = 1) \overset{d}{=} (X_s|Z_1, R = 1)\}$.*

The definition of counterfactual experiments does not specify the counterfactual target covariate $\widetilde{X}_s$ directly. It is implicitly defined through the first two elements of ϕ. The first element, $\widetilde{\mathcal{U}}_s$, is a set of rank variables associated with the counterfactual target covariate, $\widetilde{X}_s$. When $\widetilde{X}_s$ is absolutely continuous, $\widetilde{\mathcal{U}}_s$ becomes a singleton set, but the set is generally not degenerate when the distribution of $\widetilde{X}_s$ contains a mass point. The second component, $\widetilde{G}_s$, is the counterfactual distribution of $\widetilde{X}_s$ conditional on the study population. When the target covariate is continuously distributed, $\widetilde{G}_s$ is continuous and strictly increasing, and therefore, $\widetilde{X}_s$ is uniquely determined by $\widetilde{X}_s = \widetilde{G}_s^{-1}(\widetilde{U}_s)$, where $\widetilde{U}_s$ is the only element in $\widetilde{\mathcal{U}}_s$. However, when the target covariate contains mass points, there is a set of counterfactual rank variables that correspond to the same target covariate in the study population. This equivalent class is defined by Condition (i).

Following Rothe (2012), we restrict our attention to counterfactual changes where only the marginal distribution of X_s is changed, while the marginal distribution of Z and the dependence structure between X_s and Z remain unaffected. This notion of a *ceteris paribus* change is formally characterized by Conditions (ii)–(iv). Condition (ii) implies that the joint distribution

of the observed variables (Z, R) and the latent variable ϵ_s remain unchanged across counterfactual experiments. Under Condition (iii), the structural production function, g is also not affected by the counterfactual change. Condition (iv) imposes a rank similarity condition. It says the conditional rank of the counterfactual target covariate follows the same distribution as the status quo. Due to the possibility of multiplicity of rank variables, the condition is also framed in terms of a set equivalence condition. When we restrict attention to absolutely continuous target covariates, both $\mathcal{U}_s$ and $\widetilde{\mathcal{U}}_s$ are singleton sets. Hence, this condition reduces to $(\widetilde{X}_s | \widetilde{Z}_1, \widetilde{R} = 1) \stackrel{d}{=} (X_s | Z_1, R = 1)$.

Each counterfactual experiment ϕ represents a modification of the underlying economic system. It completely determines the counterfactual outcome in the study population. Yet we remain largely agnostic as to the counterfactual change in the auxiliary population. The definition also leaves the mechanism causing the change in the marginal distribution of the target covariate unspecified.

Remark 1 *Our definition of counterfactual experiments relaxes the rank invariance conditions imposed by Rothe (2012). Instead, counterfactual changes in our context only need to satisfy a rank similarity or copula invariance condition.*

With the counterfactual experiments defined, we now construct the counterfactual covariate vector by $\widetilde{W}_G := (\widetilde{G}_s^{-1}(\widetilde{U}_s), \widetilde{Z}_1')'$. The counterfactual outcome of the study population is then defined as $\widetilde{Y}_s = \widetilde{g}_s(\widetilde{W}_G, \widetilde{\epsilon}_s)$, which follows a marginal distribution, $F_{\widetilde{Y}_s}$, and a conditional distribution restricted to the principal population, $F_{\widetilde{Y}_s | R = 1}$. Note that the unconditional distribution is not well-defined, due to the lack of information on counterfactual changes in the auxiliary population. Therefore, we focus exclusively on the counterfactual distribution conditional on the study population in what follows. When X is discrete, a single counterfactual experiment is mapped to a set of counterfactual outcomes, and we denote the corresponding set of counterfactual distributions by $\mathcal{F}_{\widetilde{Y}_s}$.

In our context, the sequence of counterfactual distributions is defined in terms of the "marginal" distribution of the potential covariate X_s in the study population, rather than the true unconditional distribution of the observed X. Although X_s is missing from the main dataset, and therefore, its marginal distribution cannot be directly identified from the study population, we show in Theorem 2 that it can be recovered from the auxiliary data under the rank similarity assumption we impose in Assumption 1.

Figure 1: Marginal distributional shift and marginal quantile shift. The blue curve depicts the data distribution of X. The red curve depicts a counterfactual distribution in the sequence $\{G_t\}_{t\downarrow 0}$, which can be induced by two equivalent policy changes.

The policy parameter we seek to identify in this paper is the pathwise derivative of counterfactual distributional effect conditional on the study population. It is adapted from the definition of the *marginal partial distributional policy effect* (MPPE) by Rothe (2012).

Definition 2 (Marginal Partial Distributional Policy Effect) *Let $\Phi^* := \{\phi_t\}_{t\geq 0} \subset \Phi$ denote a sequence of counterfactual experiments, such that $\widetilde{G}_{s,t} \to F_{X_s|R=1}$, as $t \downarrow 0$. The MPPE for a given functional $\nu : \mathcal{D}(\mathcal{Y}) \to \mathbb{R}$ and a sequence of $\widetilde{F}_{s,t} \in \mathcal{F}_{\widetilde{Y}_{s,t}}$ is defined by,*

$$MPPE(\nu, \{\widetilde{Y}_{s,t}\}_{t\geq 0}) := \left.\frac{\partial \nu(F_{\widetilde{Y}_{s,t}|R=1})}{\partial t}\right|_{t=0} = \lim_{t\downarrow 0} \frac{\nu(F_{\widetilde{Y}_{s,t}|R=1}) - \nu(F_{Y_s|R=1})}{t}.$$

We consider two specific types of counterfactual distributional changes: MDS and MQS. The definition of the former is due to Firpo et al. (2009*b*). It denotes a small perturbation in the distribution of X_s, in the direction of G. MQS, on the other hand, considers a minuscule change in the quantiles of X_s. This type of policy change includes the MLS, $G_{t,ls}^{-1}(u) := F_{X_s|R}^{-1}(u|1) + t$, as a special case.

Definition 3 (Counterfactual Policy Distributions)

- *Marginal Distributional Shift (MDS): $G_{t,p}(x) := F_{X_s|R}(x|1) + t(G(x) - F_{X_s|R}(x|1))$.*

- *Marginal Quantile Shift (MQS): $G_{t,q}^{-1}(u) := F_{X_s|R}^{-1}(u|1) + t(G^{-1}(u) - F_{X_s|R}^{-1}(u|1))$.*

Remark 2 *Figure 1 illustrates how the rates of change between the two types of counterfactuals are related. Under the condition that $F_{X_s|R=1}$ is compactly supported with strictly positive density on $\mathcal{X}$, MQS in a user-specified direction, $q(x)$, can be approximated in the limit by MDS with $G(x) = F_{X_s|R}(x|1) - f_{X_s|R}(x|1)q(x)$.*

Turning to the case of quantiles, the quantile operator for a particular τ is defined by, $\nu_\tau(F_{Y_s|R=1}) := F_{Y_s|R=1}^{-1}(\tau)$. With the understanding that MPPE associated with a counterfactual experiment is generally a set when the X is discretely valued, we suppress the index with respect to $\{\widetilde{Y}_{s,t}\}_{t\geq 0}$ for notational convenience, and denote the MPPE with MDS, $MPPE(\nu_\tau, \{\widetilde{Y}_{s,t}\}_{t\geq 0})$, and MPPE with MQS, $MPPE(\nu_\tau, \{\widetilde{Y}_{s,t}\}_{t\geq 0})$, by $UQE_p(\tau, G)$ and $UQE_q(\tau, G)$, respectively. Here, and in what follows, the qualifier "unconditional" in UQE should be understood as conditional on (or relative to) the study population.

3.2 Identification of $F_{Y_s|X_s Z_1 R=1}$

If there is no missing variable, the joint distribution of (Y, X, Z) is directly identifiable from a random sample. Under data combination, however, only the "marginal" conditional distributions: $F_{Y_s|ZR=1}$ and $F_{X_a|ZR=0}$, can still be separately identified from the two samples, respectively. The conditional distribution, $F_{Y_s|X_s Z_1 R=1}$, is generally not identifiable without further cross-population assumptions.

Instead of seeking identification of the entire conditional distribution, $F_{Y_s|X_s Z_1 R=1}(\cdot|\cdot,\cdot,1)$, we demonstrate in Sections 3.3 and 3.4 that UQE, and MPPE in general, can be identified using information on a finite set of points of Y_s. Choice of the set depends on research interest. For instance, it can include only the median, the quartiles of Y_s, etc. This flexibility allows us to obtain identification under much milder restrictions on the pseudo-merged population.

Identification is achieved through the excluded instrument variables, Z_2. To ease notational burden, define $\Lambda(x, z_1) := F_{Y_s|X_s Z_1 R}(q_\tau|x, z_1, 1)$ for all $x \in \mathcal{X}$ and $z_1 \in \mathcal{Z}_1$.

Assumption 2 $\epsilon_s \perp\!\!\!\perp Z_2|X_s, Z_1, R = 1$.

Assumption 2 implies that Z_2 can be excluded from the outcome equation, and therefore, can be used as a source of exogenous variation to proxy for the missing covariate in the study population. Note that Λ is generally not identified without an exogenous instrument Z_2. We illustrate this point with a linear normal model in Example 1.

Under Assumptions 1 and 2, the following moment-matching equation holds, for all $z \in \mathcal{Z}$,

$$\mathbb{E}\left[1(Y \le q_\tau)|Z, R = 1\right] - \mathbb{E}\left[\Lambda(W)|Z, R = 0\right] = 0. \tag{3}$$

Equivalently, Λ can be identified based on a likelihood-ratio-weighting equation,

$$\mathbb{E}\left[\left. R1(Y \le q_\tau) - (1 - R)\Lambda(W)\frac{r(Z)}{1 - r(Z)} \right| Z \right] = 0. \tag{4}$$

The next assumption is about the global identification of Λ.

Assumption 3 Λ *is the unique solution to* (3) *or* (4) *almost surely.*

Assumption 3 is a high level condition. It is implied by a bounded completeness condition on the auxiliary population.[2] Note that Λ is globally identified as long as $\mathbb{E}[\Lambda(W) - \widetilde{\Lambda}(W)|Z, R = 0] = 0$ implies $\Lambda = \widetilde{\Lambda}$, which follows immediately if Λ is measurable with respect to W, and W is bounded complete for Z, relative to the auxiliary population.[3] Although Canay et al. (2013) show that the completeness condition is untestable against general alternatives, we use two examples to show that the assumption is reasonable in some special cases.

The completeness condition implicitly imposes some constraints on the support of the excluded instrument. Whenever X is continuously distributed, Z_2 is generally required to be continuous. In Example 1, we show that when both X and Z_2 are continuous, Assumption 3 holds when the structural errors follow a joint normal distribution, which is a commonly-adopted assumption in empirical practices. However, our method does not apply if the instrument has finite support or otherwise violates the bounded completeness condition, the latter of which is likely to occur if the strength of the instrument is weak.

On the other hand, when X is discretely valued, we show via Example 2 that Assumption 3 can be satisfied with a discrete instrument. The key requirement is a rank condition on conditional probability matrices of X_a given Z. When the set of Z_1 is empty, we can uses Cragg and Donald (1996) or Robin and Smith (2000) to test the rank condition.

In Section 4, we base our estimation and inference on parametric identification of Λ. In this case, we assume that $F_{Y_s|X_s Z_1 R}(q_\tau|x, z_1, 1) = \Lambda(x, z_1; \beta_0)$, for some $\beta_0 \in \theta_\beta \subset \mathbb{R}^{d_\beta}$. In Lemma A.1, we provide a set of sufficient conditions which allow us to establish a global parametric identification condition analogous to Assumption 3.

---

2 For two random element U and V, we say U is bounded complete for V, relative to a subpopulation $S = s$, if for all bounded measurable functions $\delta(\cdot)$, $\mathbb{E}[\delta(U)|V, S = s] = 0$ implies $\delta(U) \equiv 0$ almost surely.

3 Bounded completeness is weaker than the commonly adopted completeness condition appearing in Newey and Powell (2003) and Fan et al. (2014). We refer readers to Hoeffding et al. (1977), Blundell et al. (2007), and Lehmann (1986) for detailed discussions.

Lemma 1 $F_{Y_s|X_sZ_1R}(q_\tau|\cdot,\cdot,1)$ *is point identified under Assumptions 1–3.*

Lemma 1 establishes the nonparametric identification of Λ. The proof for the parametric case follows along exactly the same line so we omit it here. In the next example, we verify the identification assumptions in a conditional normal model.

Example 1 (Conditional Normal Model) *Let the structural equations of the study population be given by*

$$Y_s = g_s(X_s, Z_1) + \epsilon_s,$$

$$X_s = h_s(Z) + \eta_s,$$

where ϵ_s and η_s are jointly normally distributed. Specifically, for positive-valued functions $\psi_y(\cdot)$ and $\psi_x(\cdot)$, we have

$$(\epsilon_s, \eta_s)|Z, R = 1 \sim N\left(0, \begin{pmatrix} \psi_y(Z_1) & 0 \\ 0 & \psi_x(Z_1) \end{pmatrix}\right).$$

Then, $\Lambda(w) = \Phi(\psi_y(z_1)^{-1/2}(q_\tau - g_s(w)))$, where $\Phi(\cdot)$ denotes the CDF of standard normal distribution. Suppose the reduced-form of X given Z in the auxiliary population is $X_a = h_a(Z) + \eta_a$, Assumption 1(e) is satisfied if $h_s = h_a = h$, and $(\eta_s|Z, R = 1) \overset{d}{=} (\eta_a|Z, R = 0)$. Assumption 2 holds if $Z_2 \perp\!\!\!\perp \epsilon_s|X_s, Z_1, R = 1$. Assume, in addition that, conditional on z_1, $\text{Supp}(F_{Z_2})$ contains an open set and that $h(z_1, \cdot)$ maps open sets of z_2 into open sets. Assumption 3 then follows by Theorem 2.2 in Newey and Powell (2003).

Turning to the linear case, let $g_s(w) = \gamma_{s_1}x + \gamma'_{s_2}z_1$, $h(z) = \delta'_1 z_1 + \delta'_2 z_2$, $\psi_y = \psi_x = 1$, $\eta_s \perp\!\!\!\perp (\epsilon_s, Z_2)$, and therefore, $\mathbb{E}[1(Y \leq q_\tau)|Z, R = 1] = \Phi((q_\tau - (\gamma_{s_1}\delta'_1 + \gamma_{s_2})'Z_1 - \gamma_{s_1}\delta'_2 Z_2)/(1 + \gamma^2_{s_1})^{1/2})$. As a consequence, $(\gamma_{s_1}, \gamma'_{s_2})'$ are uniquely determined by (3) or (4), if and only if $\delta_2 \neq 0$.

Example 2 (Discrete Covariates) *Suppose X_s, X_a, and Z_2 are all discretely valued. Assume that $\text{Supp}(F_{X_j|Z_1=z_1}) = \{x^1, ..., x^l\}$ and $\text{Supp}(F_{Z_2|Z_1=z_1}) = \{z^1, ..., z^k\}$, for $j = s, a$ and for all $z_1 \in \text{Supp}(F_{Z_1})$. Let $P^j_{u,t}(z_1, r) = Prob(X_j = x^u|Z_2 = z^t, Z_1 = z_1, R = r)$ for $j = s, a$ and $r = 0, 1$. Assumption 1(e) holds if for all $u \in \{1, 2, ..., l\}$ and $t \in \{1, 2, ..., k\}$, $P^s_{u,t}(Z_1, 1) = P^a_{u,t}(Z_1, 0)$ with probability 1. Moreover, let $P^j(\cdot, \cdot)$ denote the matrix of probabilities where the (u, t)-th entry is equal to $P^j_{u,t}(\cdot, \cdot)$. Then by Theorem 2.4 in Newey and Powell (2003), bounded completeness, and hence, Assumption 3, hold if $rank(P^a(Z_1, 0)) = l$ with probability 1.*

3.3 Identification with Continuously Distributed X

In this section, we establish the identification of UQE_q and UQE_p when the distribution of X is absolutely continuous. Before stating the main result, we need some additional identifying assumptions.

Assumption 4

(a) (i) $\epsilon_s \perp\!\!\!\perp U_s | Z_1, R = 1$; (ii) there exists a t_0 sufficiently close to 0, such that for all $t \leq t_0$ and $\phi_t \in \Phi^*$, $\widetilde{\epsilon}_{s,t} \perp\!\!\!\perp \widetilde{U}_{s,t} | \widetilde{Z}_{1,t}, \widetilde{R}_t = 1$.

(b) $\mathrm{Supp}(G) \subset \mathrm{Supp}(F_{X|ZR}(\cdot|Z,1))$ almost surely.

Assumption 5 $F_{Y|R=1}$ is continuously differentiable in an open neighborhood of q_τ with strictly positive density function $f_{Y|R=1}$.

Assumption 4(a)(i) imposes that conditional on Z_1, structural error ϵ_s is independent of the rank variable U_s in the study population. This is much weaker than the commonly assumed strict independence condition that X_s is independent of ϵ_s unconditionally. Conditional exogeneity has also been imposed by Firpo et al. (2009b), Rothe (2012), and Chernozhukov, Fernández-Val and Melly (2013), among others. Assumption 4(a)(ii) requires the conditional independence condition of part (a) to hold when counterfactual experiments get sufficiently close to the status quo. Under the rank invariance condition imposed by Rothe (2012), it is automatically implied by Assumption 4(a)(i). Assumption 4(b) ensures that the conditional distribution of Y_s given W is identified over the support of W. Assumption 5 imposes a smoothness condition on the distribution of target outcome, which implies that $F_{Y|R=1}^{-1}$ is Hadamard differentiable at $F_{Y|R=1}$, tangentially to the set of functions that are continuous at q_τ.

The main theoretical result of this section is given as follows.

Theorem 2 *Suppose that Assumptions 1–5 hold, and that the distribution of X is absolutely continuous with respect to the Lebesgue measure, both $UQE_p(\tau, G)$ and $UQE_q(\tau, G)$ are identified.*

(a) For UQE_q, we have

$$UQE_q(\tau, G) = -\frac{1}{f_{Y_s|R}(q_\tau|1)(1 - Q_0)}\mathbb{E}\left[(1 - R)\ell(Z)\Lambda_x(X, Z_1)g_q(X)\right],$$

where $g_q(x) := G^{-1}(F_{X_s|R}(x|1)) - x$, and $F_{X_s|R}(x|1) = \frac{1}{1-Q_0}\mathbb{E}\left[(1 - R)\ell(Z)1(X \leq x)\right]$.

(b) Suppose in addition that $\mathcal{X}$ is compact, and $F_{X_s|R=1}$ is continuously differentiable on $\mathcal{X}$ with strictly positive density function $f_{X_s|R=1}$. Then we have,

$$UQE_p(\tau, G) = -\frac{1}{f_{Y_s|R}(q_\tau|1)(1-Q_0)}\mathbb{E}\left[(1-R)\ell(Z)\Lambda_x(X, Z_1)g_p(X)\right],$$

where $g_p(x) := -\frac{G(x)-F_{X_s|R}(x|1)}{f_{X_s|R}(x|1)}$, and $f_{X_s|R}(x|1) = \frac{1}{1-Q_0}\partial\mathbb{E}\left[(1-R)\ell(Z)1(X \le x)\right]/\partial x$.

Remark 3 *The compactness condition on $\mathcal{X}$ is assumed to ensure the existence of pathwise derivative of the inverse map. It can be relaxed by imposing a boundary condition on Λ_x. Specifically, we may assume that Λ_x vanishes when $x \notin [F_{X_s|R=1}(q_1) + \epsilon, F_{X_s|R=1}(q_2) - \epsilon]$, for $0 < q_1 < q_2 < 1$ and some $\epsilon > 0$.*

3.4 Identification with Discrete Covariate

Let the support of X be $\{x^1, \ldots, x^l\}$. When X is discrete, MQS is not well-defined and we consider MDS only, with counterfactual experiments defined through a fixed discrete distribution, G. As indicated by Example 2, results in this section hold when Z_2 are both continuously and discretely valued.

Assumption 6

(a) (i) $\epsilon_s \perp\!\!\!\perp U_s|Z_1, R = 1$, for all $U_s \in \mathcal{U}_s$; (ii) there exists a t_0 sufficiently close to 0, such that for all $t \le t_0$ and $\phi_t \in \Phi^*$, $\epsilon_{s,t} \perp\!\!\!\perp \widetilde{U}_{s,t}|\widetilde{Z}_{1,t}, \widetilde{R}_t = 1$, for all $\widetilde{U}_{s,t} \in \widetilde{\mathcal{U}}_{s,t}$.

(b) $\text{Supp}(G) \subset \text{Supp}(F_{X_s|R=1})$.

(c) For all $U_s \in \mathcal{U}_s$, $F_{U_s|Z_1R}(u_s|z_1, 1)$ is continuously differentiable in u_s, for all $z_1 \in \mathcal{Z}_1$.

Assumption 6(a) is the counterpart of Assumption 4(a) for discrete covariates. Since the rank variables are no longer uniquely pinned down by strictly increasing quantile functions, we strengthen Assumption 4(a) so that conditional independence holds for all the rank variables in the equivalent class. With this identifying assumption in hand, we are ready to present the following identification result. For $j = 1, \ldots, l$, let the period bound generating function be defined by

$$h_{q_\tau}(x^j, x^{j-1}, z_1) := -\frac{(\Lambda(x^{j-1}, z_1) - \Lambda(x^j, z_1)) \cdot (G(x^{j-1}) - F_{X_s|R}(x^{j-1}|1))}{f_{Y_s|R}(q_\tau|1)}.$$

Theorem 3 *Suppose that Assumptions 1–3, 5, and 6 hold, $UQE_p(\tau, G)$ is partially identified, with*

$$UQE_p(\tau, G) \in \left[\sum_{j \in \mathcal{J}_+} h_{q_\tau}(x^j, x^{j-1}, z_{1,j}^\dagger) + \sum_{j \in \mathcal{J}_-} h_{q_\tau}(x^j, x^{j-1}, z_{1,j}^*), \right.$$

$$\left. \sum_{j \in \mathcal{J}_+} h_{q_\tau}(x^j, x^{j-1}, z_{1,j}^*) + \sum_{j \in \mathcal{J}_-} h_{q_\tau}(x^j, x^{j-1}, z_{1,j}^\dagger) \right],$$

where $\mathcal{J}_+ := \{j \in \{1, \ldots, l\} : G(x^{j-1}) \leq F_{X_s|R}(x^{j-1}|1)\}$ ($\mathcal{J}_-$ is analogously defined), $z_{1,j}^* := \arg\sup_{z_1 \in \mathcal{Z}_1}(\Lambda(x^{j-1}, z_1) - \Lambda(x^j, z_1))$, $z_{1,j}^\dagger := \arg\inf_{z_1 \in \mathcal{Z}_1}(\Lambda(x^{j-1}, z_1) - \Lambda(x^j, z_1))$, and $F_{X_s|R}(x^j|1) = \mathbb{E}[\frac{1-R}{1-Q_0}\ell(Z)1(X \leq x^j)]$, for $j \in \{1, \ldots, l\}$.

Theorem 3 indicates that UQE_p is generally partially identified with bounds generated by h_{q_τ}. In the special case when $\Lambda(x, z_1)$ is constant in z_1, the identified set of UQE_p reduces to a singleton.

If X is binary and $G_{t,p}(x) = 1\{0 \leq x < 1\}(F_{X_s|R}(0|1) - t) + 1\{x \geq 1\}$, h_{q_τ} reduces to $-(\Lambda(1, z_1) - \Lambda(0, z_1))/f_{Y|R}(q_\tau|1)$. In such circumstance, Theorem 3 corresponds to the two-sample generalization of Theorem 5 in Rothe (2012), when ν in that paper takes on the quantile functional.

4 Estimation and Inference

In this section, we discuss estimation and inference for our two-sample UQE. First, we describe an estimation procedure for UQE_q and UQE_p as identified in Theorem 2. We then show that our estimator is consistent and asymptotically normal in Theorem 5.[4]

4.1 Estimation Procedure

Following the discussion in Section 3, we first propose an estimator of the conditional probability, Λ. Here, we restrict our attention to the parametric setting where Λ is indexed by a vector of parameter, β. We use a moment-matching method based on (4) to estimate β. The estimation of β consists of four-steps. In the first step, we estimate q_τ by solving

$$\widehat{q}_\tau := \arg\min_{q \in \mathcal{Y}} \mathbb{E}_n[R(\tau - 1(Y \leq q)) \cdot (Y - q)]. \tag{5}$$

The next three steps follow closely the Auxilliary-to-Study Tilting (AST) method proposed by Graham et al. (2016). Using the AST estimator, β and the propensity score can be jointly

---

4 Here we focus on the scenario where the distribution of X is absolutely continuous. When X is discrete, the problem features partially identified parameters defined by the intersection bounds. Chernozhukov, Lee and Rosen (2013) provide an extensive treatment of this topic. We omit discussion here and refer readers to Appendix D in Rothe (2012) for a detailed discussion on how to apply their method.

estimated from moment restrictions in (4). To implement the estimator, we first estimate the propensity score, $r(z)$. Towards this end, we assume that the propensity score takes a parametric form, i.e. $r(z) = L(k(z)'\gamma)$, where $L(\cdot)$ is any link function that satisfies Assumption 7(e). Using $L(\cdot)$, $\widehat{\gamma}$ can be obtained by solving the following problem,

$$\widehat{\gamma} := \arg\max_{\gamma \in \Theta_\gamma} \mathbb{E}_n \left[R \log(L(k(Z)'\gamma)) + (1-R)\log(1 - L(k(Z)'\gamma)) \right]. \tag{6}$$

The AST estimator augments the conditional maximum likelihood estimator $\widehat{\gamma}$ with tilting parameters. The resulting estimator of β is more efficient than the one based on $\widehat{\gamma}$ alone. Let $t(z)$ be a vector of known functions of z with a constant term as the first element. Denote the tilting parameters associated with the auxiliary data and the study sample, by λ_a and λ_s, respectively. They are estimated by solving,

$$\mathbb{E}_n \left[\left(\frac{1-R}{1 - L(k(Z)'\widehat{\gamma} + t(Z)'\widehat{\lambda}_a)} - 1 \right) L(k(Z)'\widehat{\gamma})t(Z) \right] = 0, \tag{7}$$

$$\mathbb{E}_n \left[\left(\frac{R}{L(k(Z)'\widehat{\gamma} + t(Z)'\widehat{\lambda}_s)} - 1 \right) L(k(Z)'\widehat{\gamma})t(Z) \right] = 0. \tag{8}$$

Using $\widehat{\lambda}_s$ and $\widehat{\lambda}_a$, we compute study and auxiliary sample tilts, which are defined as follows

$$\widehat{\pi}_i^s := \frac{L(k(Z_i)'\widehat{\gamma})}{L(k(Z_i)'\widehat{\gamma} + t(Z_i)'\widehat{\lambda}_s)}, \qquad \widehat{\pi}_i^a := \frac{L(k(Z_i)'\widehat{\gamma})}{1 - L(k(Z_i)'\widehat{\gamma} + t(Z_i)'\widehat{\lambda}_a)}. \tag{9}$$

Also let $e(z)$ be a d_β-dimensional vector of known functions of z, and $g(a; \widehat{q}_\tau, \widehat{\gamma}, \widehat{\lambda}_s, \widehat{\lambda}_a, \beta) := (\widehat{\pi}^s r 1(y \leq \widehat{q}_\tau) - \widehat{\pi}^a(1-r)\Lambda(w;\beta))e(z)$. Now, in the last step, β can be estimated by

$$\widehat{\beta} := \arg\inf_{\beta \in \Theta_\beta} \widehat{\mathcal{L}}_n(\beta), \tag{10}$$

where $\widehat{\mathcal{L}}_n(\beta) := \left\| \mathbb{E}_n[g(A; \widehat{q}_\tau, \widehat{\gamma}, \widehat{\lambda}_s, \widehat{\lambda}_a, \beta)] \right\|_{\Omega_n}^2$ and $\|x\|_{\Omega_n}^2 := x'\Omega_n x$, for a sequence of positive definite weighting matrices Ω_n.

Using these quantities, we can obtain $\Lambda_x(W;\widehat{\beta}) := \partial\Lambda(W;\widehat{\beta})/\partial x$, and $\widehat{\ell}(z) := \frac{n_a}{n_s} \cdot \frac{L(k(z)'\widehat{\gamma}+t(z)'\widehat{\lambda}_s)}{1 - L(k(z)'\widehat{\gamma}+t(z)'\widehat{\lambda}_a)}$. Throughout this section, we assume that the counterfactual distribution G is known. In practice, if G is not known, it may be estimated from an independent sample; see e.g. Rothe (2010). Using the above estimates and $\widehat{F}_{X_s|R=1}(\cdot) := \mathbb{E}_{n_a}[\widehat{\ell}(z)1(X \leq \cdot)]$, where $\mathbb{E}_{n_a}[X]$ denotes $n_a^{-1} \sum_{i=n_s+1}^n X_i$, $\widehat{g}_q$ can be obtained as the plug-in estimator. For g_p, we need an estimator for $f_{X|R=1}(\cdot)$. Our identification relies on a compact support condition, and it is well known that the Prazen-Rosenblatt density estimator is not valid near the boundary of support. To overcome this challenge, we introduce trimming.[5] For a kernel K_x with compact support, and

---

[5] Trimming is widely adopted in the literature; see e.g. Härdle and Stoker (1989), Powell et al. (1989) among others. This specific trimming function is inspired by Guerre et al. (2000) and Li et al. (2002). As an alternative,

some bandwidth b_x, we let

$$\widehat{f}_{X|R}(x|1) := \mathbb{E}_{n_a}\left[\widehat{\ell}(Z)I_{b_x}K_{b_x}(X-x)\right],$$

where $K_{b_x}(\cdot) := b_x^{-1}K_x(\cdot/b_x)$. I_{b_x} is a trimming indicator, which equals one for $x \in \{[\underline{x} + \rho_x b_x/2, \bar{x} - \rho_x b_x/2]\}$, where $\underline{x}$, $\bar{x}$, and ρ_x are the lower and upper bound, of $\mathcal{X}$, and the diameter of $\text{Supp}(K_x)$, respectively. The density, $f_{Y|R}$, can also be estimated using a kernel density estimator. Specifically, for any kernel function $K_y(\cdot)$ that satisfies Assumption 8(b), let $\widehat{f}_{Y|R}(y|1) := \mathbb{E}_{n_s}[K_{b_y}(Y_i - y)]$, where $\mathbb{E}_{n_s}[X] := n_s^{-1}\sum_{i=1}^{n_s} X_i$ and $K_{b_y}(y) := b_y^{-1}K_y(y/b_y)$.

Now, plugging in the estimators of nuisance quantities, $UQE_j(\tau, G)$ can thus be estimated by,

$$\widehat{UQE}_j(\tau, G) := -\frac{1}{\widehat{f}_{YR}(\widehat{q}_\tau|1)}\mathbb{E}_{n_a}[\widehat{\ell}(Z)\Lambda_x(W;\widehat{\beta})\widehat{g}_j(X)], \ \ j = p, q. \tag{11}$$

We summarize the estimation procedure in the following algorithm.

Algorithm 1 (Plug-in Estimator for $\widehat{UQE}$)

1. *Compute the empirical quantile estimator $\widehat{q}_\tau$ by solving (5).*

2. *Compute the conditional maximum likelihood estimator $\widehat{\gamma}$ by solving (6).*

3. *Solve (7) and (8) to get $\widehat{\lambda}_j$, and use them to compute $\widehat{\pi}_j$, for $j = s, a$, following (9).*

4. *Use the above quantities to compute $\widehat{\beta}$, by solving (10).*

5. *Compute $\Lambda_x(\cdot;\widehat{\beta}), \widehat{\ell}(\cdot), \widehat{F}_{X_s|R=1}(\cdot), \widehat{f}_{X_s|R=1}(\cdot)$. Using these quantities to compute $\widehat{g}_j$, for $j = p, q$.*

6. *For $j = p, q$, compute the plug-in estimator $\widehat{UQE}_j$ following (11).*

4.2 Large Sample Results

In this section, we present inference results for the estimators introduced in the previous section. We first establish large sample properties of $\widehat{\beta}$, for which purpose, some additional regularity conditions are in order.

Assumption 7

---

we can use a local polynomial density estimator that adjusts for the boundary bias adaptively; see Cattaneo et al. (2020) for details.

(a) (i)$\{(R_i, R_iY_i, (1-R_i)X_i, Z_i)\}_{i=1}^n$ are i.i.d.; (ii) let $\theta := (\gamma, \lambda_s, \lambda_a, \beta) \in \Theta := \Theta_\beta \times \Theta_\lambda^2 \times \Theta_\beta$, then Θ is compact, and θ_0 lies in the interior of Θ.

(b) $F_{Y|ZR=1}(y|z)$ is absolutely continuous and differentiable in $y \in \mathcal{Y}_0$ for all $z \in \mathcal{Z}$, where $\mathcal{Y}_0$ is a compact subset of $\mathcal{Y}$, and

$$\sup_{(y,z)\in\mathcal{Y}_0\mathcal{Z}} |f_{Y|ZR}(y|z,1)| \le c_1 < \infty.$$

(c) (i) $\Lambda(w;\beta)$ is twice continuously differentiable in β with uniformly bounded derivatives, for all $w \in \mathcal{W}$; (ii) $0 \le \inf_{w,\beta} \Lambda(w;\beta), \sup_{w,\beta} \Lambda(w;\beta) \le 1$; (iii) $\Lambda_x(\cdot;\beta)$ is continuously differentiable in β, and $\sup_{w,\beta} |\Lambda_x(w;\beta)| \le c_2 < \infty$.

(d) There exists a symmetric, non-random matrix Ω, such that $||\Omega_n - \Omega|| = O_p(\delta_{\omega,n})$, where $\delta_{\omega,n} = o(1)$, and that $c_3^{-1} \le \lambda_{min}(\Omega) \le \lambda_{max}(\Omega) \le c_3$.

(e) There is a unique $\gamma_0 \in \Theta_\gamma$, and known function $L(\cdot)$ such that (i)

$$\ell(z) = \frac{1-Q_0}{Q_0} \cdot \frac{L(k(z)'\gamma_0)}{1-L(k(z)'\gamma_0)}.$$

(ii) $L(\cdot)$ is strictly increasing, twice continuously differentiable, with bounded first and second order derivatives; (ii) $\lim_{x\to-\infty} L(x) = 0$ and $\lim_{x\to\infty} L(x) = 1$; (iii) $0 < c_4 < L(k(z)'\gamma + t(z)'\lambda_j) \le c_5 < 1$ for all $(\gamma, \lambda_j) \in \Theta_\gamma \times \Theta_\lambda$, $j = s, a$, and $z \in \mathcal{Z}$.

(f) $\mathbb{E}[||j(Z)||^4] < \infty$, where $j = k, t, e$.

Assumption 7(a) is standard in the microeconometric literature. Assumption 7(b) requires the conditional density $f_{Y|ZR}(\cdot|\cdot, 1)$ be bounded uniformly for all $(y, z) \in \mathcal{Y}_0 \mathcal{Z}$. Assumption 7(c) imposes mild smoothness conditions on the parametric function $\Lambda(\cdot, \cdot; \cdot)$, requiring it to be bounded between the unit interval, thus behaving like a distribution function. Assumption 7(d) states that Ω_n is consistent for Ω, which is positive definite. Assumption 7(e) implies that the true "propensity score" is known up to finite dimensional γ_0. It also specifies smoothness and boundedness conditions on the parametric propensity score. Finally, due to the estimation of q_τ, we impose a finite fourth moment condition in Assumption 7(f), which is stronger than the usual square-integrability condition.

Lemma 4 Suppose that Assumptions 1–5 and Assumption 7 hold, then (i) $\widehat{\beta} \xrightarrow{p} \beta_0$; furthermore, (ii) suppose that the Jacobian, M_Ω, as defined in (A.5) in the Online Appendix, is in-

vertible, then

$$\sqrt{n}(\widehat{\beta} - \beta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_\beta(A_i; \theta_0, q_\tau) + o_p(1),$$

where $\psi_\beta(A; \theta_0, q_\tau)$ is given by (A.8) in the Online Appendix, and (iii)

$$\sqrt{n}(\widehat{\beta} - \beta_0) \xrightarrow{d} N(0, \Sigma_\beta),$$

where $\Sigma_\beta := \mathbb{E}[\psi_\beta(A; \theta_0, q_\tau)\psi_\beta(A; \theta_0, q_\tau)']$.

Lemma 4 shows that the parameters of $F_{Y_s|X_sZR=1}$ are consistently estimated by $\widehat{\beta}$. Furthermore, it admits an asymptotic linear representation with influence function given by $\psi_\beta(A; \theta_0, q_\tau)$, which plays a key role in establishing the large sample properties of UQE. Towards this ends, we need the following set of assumptions.

Assumption 8

(a) (i) $F_{Y|R=1}(\cdot)$ is absolutely continuous and differentiable over $y \in \mathcal{Y}$; (ii) $f_{Y|R=1}(\cdot)$ is uniformly continuous; (iii) the density $f_{Y|R=1}(y)$ is strictly bounded away from 0, three times continuously differentiable in y with uniformly bounded derivatives for y in $\mathcal{Y}_0$, such that $q_\tau \in \mathcal{Y}_0$.

(b) The kernel function $K_y(\cdot)$ is symmetric, continuous, bounded, with a compact support, and such that (i) $\int K_y(y)dy = 1$; (ii) $\int yK_y(y)dy = 0$.

(c) $b_y \to 0$, $\log(n)n^{-1}b_y^{-1} \to 0$, and $nb_y^5 \to c_6 < \infty$.

(d) (i) $\mathcal{X}$ is compact; (ii) G is continuously differentiable on $\mathcal{X}$ with strictly positive density.

Assumption 8(a) strengthens Assumption 5 and imposes stronger smoothness conditions on the distribution of Y_s. Assumption 8(b) states several regularity conditions on kernel functions, which is standard in the literature. Assumption 8(c) specifies admissible rate for the bandwidth parameter. We can choose $b_y = O(n_s^{-\kappa})$, for $\kappa \in [1/5, 1/2)$. Assumption 8(d) imposes support and smoothness conditions for the counterfactual target covariate.

Asymptotic properties of $\widehat{UQE}$ are formally characterized in the next theorem.

Theorem 5 *Under Assumptions 1–5, 7, and 8, (i) the following linear expansions hold,*

$$\widehat{UQE}_q(\tau, G) - UQE_q(\tau, G) = \frac{1}{n} \sum_{i=1}^{n} \psi_q + B_q(\tau, G, b_y) + o_p(n^{-1/2}).$$

Suppose in addition that Assumption B.1 holds, (ii) then we have

$$\widehat{UQE}_p(\tau, G) - UQE_p(\tau, G) = \frac{1}{n}\sum_{i=1}^{n} \psi_p + B_p(\tau, G, b_y) + o_p(n^{-1/2}),$$

where, ψ_j, $j = p, q$, is defined in Appendix B, $B_j(q_\tau, G, b_y) := \frac{b_y^2 f''_{Y|R}(q_\tau|1) d_j(\theta_0, G)}{2 f^2_{Y|R}(q_\tau|1)} \cdot \int y^2 K_y(y) dy$,

and $d_j(\theta_0, G) := \frac{1}{1-Q_0}\mathbb{E}[(1-R)\ell(Z)\Lambda_x(X, Z_1; \beta_0)g_j(X)]$, for $j = p, q$.

(iii) Therefore,

$$\sqrt{nb_y}(\widehat{UQE}_j(\tau, G) - UQE_j(\tau, G) - B_j(q_\tau, G, b_y)) \xrightarrow{d} N(0, \Sigma_j),$$

where, $\Sigma_j := \frac{d_j^2(\theta_0, G)}{f^3_{Y|R}(q_\tau|1)Q_0}\int K_y^2(y)dy$, for $j = p, q$.

From the linear expansions in Theorem 5, we conclude that UQE converges at a rate that is slower than root-n. This result is mainly driven by the nonparametric estimation of the density $f_{Y|R=1}$, and therefore, the estimator is nonparametric in essence. Moreover, the asymptotic expansion includes an asymptotic bias term, $B(\tau, G, b_y)$. If we assume, as in Firpo et al. (2009a), $nb_y^5 \to 0$ or $\kappa < 1/5$, the bias vanishes asymptotically.

Remark 4 *Estimators for the asymptotic variance of $UQE_p(\tau, G)$ and $UQE_q(\tau, G)$ can be constructed using their empirical counterparts. Specifically, let*

$$\widehat{\Sigma}_j := \frac{\widehat{d}_{j,n}(\widehat{\theta}, G)^2}{\widehat{f}^3_{Y|R}(\widehat{q}_\tau|1)\,\mathbb{E}_n[R]}\int K_y^2(y)dy,$$

where $\widehat{d}_{j,n}(\widehat{\theta}, G) := \mathbb{E}_{n_a}[\widehat{\ell}(Z)\Lambda_x(W; \widehat{\beta})\widehat{g}_j(X)]$. Under a suitable rate condition on b_y, consistency of $\widehat{\Sigma}_j$ follows directly from the first two parts of Theorem 5. To achieve better finite-sample performance, we can add the root-n terms of the influence functions to the variance estimator, based on which, we propose the following improved variance estimator,

$$\widehat{\Sigma}_{j,imp} := b_y\,\mathbb{E}_n[\widehat{\psi}_j(A; \widehat{\theta}, \widehat{q}_\tau, b_y)^2]. \tag{12}$$

In the above definition, $\widehat{\psi}_j(a; \widehat{\theta}, \widehat{q}_\tau, b_y)$ is a plug-in estimator of the influence function, $\psi_j(A; \theta_0, q_\tau, b_y)$, for $j = p, q$. A detailed description of the construction of $\widehat{\psi}$ can be found in the Online Appendix C.

Remark 5 *Theorem 5 implies that tests of the unconditional quantile effect converges at a non-parametric rate in general. Nonetheless, for the null of zero, positive, and negative effects, we can still construct tests that have power against departures of the null at the parametric rate. For example, to test the null: $H_0 : UQE_j(\tau, G) = 0$, it is equivalent to test $H'_0 : d_j(\beta_0, G) = 0$, as $UQE_j(\tau, G) = 0 \Leftrightarrow d_j(\beta_0, G) = 0$, for $j = p, q$. From Theorem 5, we know that $\widehat{d}_{j,n}(\widehat{\beta}, G)$*

converges at the parametric rate. Moreover, we have

$$\sqrt{n}\widehat{V}_{d,j}^{-1/2}(\widehat{d}_{j,n}(\widehat{\theta}, G) - d_j(\theta_0, G)) \xrightarrow{d} N(0,1), \tag{13}$$

where $\widehat{V}_{d,j}$ is an estimator of $V_{d,j} := \mathbb{E}[\psi_{d,j}(A; \theta_0, q_\tau)^2]$, with $\psi_{d,j}$, $j = p, q$, defined in Appendix B. The result in (13) can be used to test H_0', applying standard testing procedures.

5   Empirical Illustration

We apply our identification and estimation methods to a variant of the Mincer's regression. Our main goal here is to demonstrate the bias from using potential instead of actual labor experience in human capital earnings models.

Identifying the causal relationship between earnings and human capital accumulation has been a focus of labor economic studies for decades. Traditionally, Mincer's regression has been widely used to quantify the link between labor wage, education and labor market experience.

Most datasets do not provide respondents' actual work histories. Therefore, many researchers choose to proxy the variable with potential work experience. The potential experience measure is usually calculated by subtracting years of schooling plus some constant (typically 6 years) from age. Despite the popularity of this practice, many labor economists believe that the return to actual experience tends to be biased when we employ the potential experience as proxy; see e.g. Regan and Oaxaca (2009). One of their main arguments is that any lapse in labor force participation would be implicitly assumed away when potential experience instead of the actual one is used. There is little reason to believe that the return to employed experience is the same as that of the unemployed period. Hence, it is still preferable to use the actual labor experience.

We use the 1970 wave of IPUMS as our main sample. The data is a 1-in-10,000 national random sample of the population. The outcome of interest is the natural log of yearly earnings. The target covariate, actual work experience, is missing from IPUMS. To apply the procedure described in Section 4, we need a dataset where the actual work experience is available. For that purpose, we use the 1972 wave of PSID as cleaned by Hirukawa et al. (2020). Detailed work histories are available in PSID. Therefore, it allows us to recover the actual labor market experience. However, running analysis directly with PSID may not be ideal due to the fact that it is not nationally representative. Our method is able to address this issue by combining information from both samples.

To estimate $F_{Y_s|X_s Z_1 R=1}$, we consider the following specification,

$$\mathbb{P}(log(Income) \leq y) = \Lambda(\beta_0 + \beta_1 educ + \beta_2 black + \beta_3 south$$

$$+ \beta_4 married + \beta_5 exper_r + \beta_6 exper_r^2) \qquad (14)$$

where $exper_r$ stands for individual's actual or realized work experience, $educ$ denotes the highest grade completed by the respondent, $black, married$, and $south$ are dummy variables which take one if the person is black, married, and lives in the south, respectively. The actual work experience serves as our X_s. It enters (14) with linear and quadratic terms. We let $(educ, black, south, married)$ be the set of included instruments Z_1, and the potential experience, $exper_p$, be the excluded instrument Z_2.

Table 1: Summary statistics

Variable	Mean	St. Dev.	Min	25th Pctl.	Median	75th Pctl.	Max
Data Source A: The IPUMS Sample (1970s)							
Income	7,923.98	6,218.37	50	4,050	7,050	10,050	50,000
Log(Income)	8.64	0.97	3.91	8.31	8.86	9.22	10.82
Age	38.29	13.88	17	26	37	50	65
Education	11.43	2.71	5	10	12	13	17
Black	0.07	0.25	0	0	0	0	1
South	0.26	0.44	0	0	0	1	1
Married	0.75	0.43	0	0	1	1	1
Potential Experience	20.86	14.7	0	8	20	33	53
Data Source B: The PSID Sample							
Income	8,966.52	5,905.47	50	5,069	8,000	11,359	70,000
Log(Income)	8.88	0.74	3.91	8.53	8.99	9.34	11.16
Age	37.8	12.37	17	26	37	47	65
Education	12.14	3.06	5	11	12	16	17
Black	0.27	0.44	0	0	0	1	1
South	0.41	0.49	0	0	0	1	1
Married	0.89	0.31	0	1	1	1	1
Potential Experience	19.65	13.42	0	7	18	30	53
Actual Experience	18.87	12.23	0	8	18	28	56
Data Source C: The IPUMS Sample (1980s)							
Potential Experience	18.99	14.49	0	6	16	31	53

Notes: Summary statistics for IPUMS and PSID. The top panel uses male subsample (aged between 17 to 65) from the 1970 wave of IPUMS with a sample size of 5,807. The middle panel uses the male subsample (aged between 17 to 65) from the 1972 wave of PSID with a sample size of 2,339. The bottom panel uses male subsample (aged between 17 to 65) from the 1980 wave of IPUMS with a sample size of 533,517.

Mincer et al. (1974) derives the relationship between schooling, labor market experience and earnings by means of an accounting identity model. We assume that realized labor experience,

rather than potential experience as constructed by econometricians, determines post-school investment, and therefore, observed earnings. This belief is embodied in Assumption 2, which requires that earnings are independent with potential experience, conditional on actual experience and education. This restriction holds if the mechanism that governs the discrepancy between potential and actual experience is unrelated to the wage determination process. Note that it in principle rules out cases where individuals leave labor market in response to wage rate fluctuations.

Visual check of the actual-experience-specific age-income profiles can serve as a preliminary test of the exclusion restriction. There are many reasons why such a parsimonious model is often refuted by data. See Heckman et al. (2006), Lemieux (2006), and the references therein for a detailed discussion of the empirics. Nevertheless, we believe that our modification of the benchmark Mincer regression suffices for an illustrative purpose.

Due to the relatively large support, we treat experience as a continuous variable. We assume that, given potential experience and the set of controls, actual labor experience follows the same distribution in the two samples, which implies that Assumption 1(e) holds. This assumption is reasonable in our context since we the two surveys were taken around the same time in history. Additionally, we assume the errors of structural equations follow a joint normal distribution. Therefore, Assumption 3 follows by Example 1.

We provide estimation results when $\Lambda(\cdot)$ takes either the logistic link or the probit link. To implement the AST estimator, we choose $j(Z) = (Z_1', Z_2')'$, for $j = k, t, e$. The density of Y_s is estimated using $b_y = n_s^{-0.01} b_{n,0}$, where $b_{n,0} := 1.06 \min\{\sigma(Y_s), interquartile(Y_s)\} n_1^{-0.2}$ is the usual "rule-of-thumb" bandwidth.

Table 1 reports the descriptive statistics for the two samples. Following the standard practice in labor economics, we use only the data on men aged between 17 and 65 when the surveys are taken. To ensure that Assumption 1(a) holds, we further trim the IPUMS sample to match the sample bounds observed in the PSID data set. This leaves us with a sample of $n_s = 5,807$ respondents for IPUMS and $n_a = 2,339$ for PSID.There are considerable differences between the two datasets. Individuals who are black, married and/or lives in the south are over-represented in PSID compared to the nationally representative IPUMS. On average, an individual in PSID has 0.78 years more potential experience than actual experience.

For UQE with MDS and MQS, we take the smoothed empirical distribution of $exper_p$ from the 1980 wave of IPUMS 1-in-100 sample (trimmed to match the support of the PSID sample)

Table 2: Estimation results

Quantile Level	0.25	0.5	0.75	0.25	0.5	0.75
		Logit Link			Probit Link	
MDS						
$UQE_{2s}(\tau)$	-0.2297	-0.0612	-0.0296	-0.2304	-0.0619	-0.0299
	(0.0587)	(0.0125)	(0.0054)	(0.0576)	(0.0124)	(0.0055)
$H_0 : UQE_{2s} = 0$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
MQS						
$UQE_{2s}(\tau)$	-0.2159	-0.0611	-0.0305	-0.2159	-0.0615	-0.0308
	(0.0779)	(0.0164)	(0.0063)	(0.0778)	(0.0164)	(0.0063)
$H_0 : UQE_{2s} = 0$	0.0043	0.0002	0.0000	0.0043	0.0002	0.0000
MLS						
$UQE_{2s}(\tau)$	0.0956	0.0241	0.0135	0.0957	0.0243	0.0143
	(0.0154)	(0.0020)	(0.0011)	(0.0151)	(0.0020)	(0.0011)
$H_0 : UQE_{2s} = 0$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Notes: In each panel, the first two rows report point estimates and standard error using our two sample estimator. The last row of each panel reports the p-value associated with the Wald test of zero effect.

as the target counterfactual distribution. The policy question we would like to answer with this counterfactual is as follows: What is the unconditional quantile effect if the distribution of labor market experience shifts marginally towards that is observed in the 1980s. Due to the large sample size of the counterfactual sample ($n = 533,517$), we can ignore the sampling variation and treat the target distribution as known. As shown in Table 1, we find that less-experienced workers tend to have even fewer years of experience in the counterfactual scenario than in the status quo, and the opposite is true for workers closer to the right tail of the distribution.

We report estimation results in Table 2 and Figure 2. A few remarks are in order. First, our estimates suggest that the counterfactual effect of a marginal shift in the distribution of actual experience is heterogeneous across income groups. The effect is larger in magnitude for the lower-income groups as expected. When MDS and MQS are considered, the quantile effects are uniformly negative and the shapes of the effect curves are similar. The marginal shift could decrease the (log) earning by anything between 0.03 and 0.23 across income quantiles. For reference, when the logistic link is assumed, the marginal effect of MDS amounts to a reduction of 20.5% in annual earnings for individuals the first quartile, 5.9% at the median, and 2.9% at the third quartile, respectively. The MLS estimates are of different signs from MDS and MQS. The marginal upward shift in the actual experience would increase a median worker's income by 2.4%.
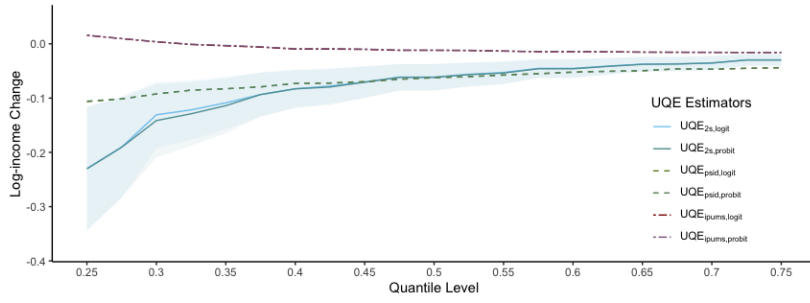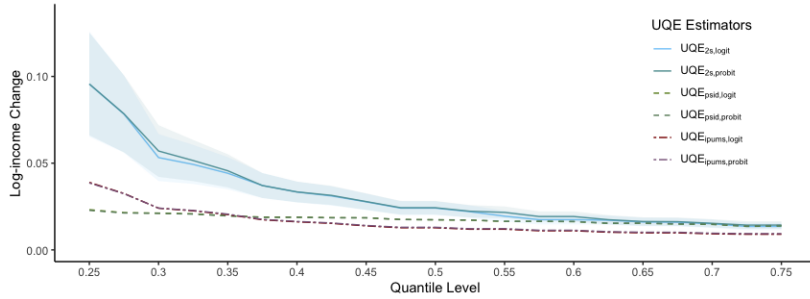
(a) MDS
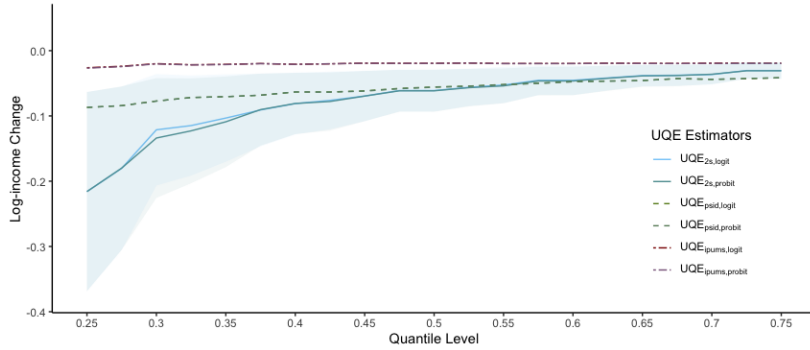
(b) MQS

(c) MLS

Figure 2: Unconditional quantile effect of actual experience on log(Earnings). The top panel: Results for UQE with MDS. The middle panel: Results for UQE with MQS. The bottom panel: Results for UQE with MLS. All three plots contain the UQE of potential experience based on IPUMS (two-dash lines), the UQE of actual experience based on PSID (dashed lines), the two-sample UQE (solid lines), and the two-sided 95% confidence intervals based on the improved variance estimator (shaded area).

Next, we consider the bias caused by using potential experience in lieu of the actual experience. We note that, one-sample UQE estimates based on IPUMS tend to be smaller in magnitude at lower income quantiles than that based on the combined data. Eventually, the two estimators converge at higher income levels.[6]

---

6 Our analysis is local to the direction of counterfactual change, and therefore, does not allow the result to be extrapolated globally. For the same reason, the comparison between UQE_{2s} and UQE_{psid} is not meaningful.

6 Concluding Remarks

In this paper, we propose a framework to identify and estimate unconditional quantile policy effect under data combination. We establish the identification of UQE under two main conditions: a rank similarity assumption and a conditional independence assumption, based on which, we provide estimators for the identified UQE and derive their large sample properties.

Our current approach can be extended in the following directions. First, although we have restricted our attention to the quantile effect throughout this paper, our results can be easily extended to other statistical functionals such as mean, interquartile, and inequality measures. It would be interesting to see how the identification requirements change with respect to the functional we adopt. Second, we have focused exclusively on the pointwise identification and inference. While extension to uniform results seem straightforward, it comes at a cost of stronger cross-sample restrictions. Under such assumptions, conditional quantile regression is likely feasible. Comparing conditional and unconditional quantile effects, as in Firpo et al. (2009*b*), under our two-sample structure, would also be an interesting direction for future research.

Appendix

A   Proofs of Lemmas and Theorems in Section 3

Proof of Lemma 1:   We provide proof for the nonparametric identification here. The proof for the parametric case follows along exactly the same line and is omitted. We shall show (3) first,

$$
\mathbb{E}[1(Y \leq q_\tau)|Z, R = 1] = \mathbb{E}[\mathbb{E}[1(Y_s \leq q_\tau)|X_s, Z, R = 1]|Z, R = 1]
$$

$$
= \mathbb{E}[\mathbb{E}[1(Y_s \leq q_\tau)|X_s, Z_1, R = 1]|Z, R = 1]
$$

$$
= \mathbb{E}[\Lambda(X_s, Z_1)|Z, R = 1]
$$

$$
= \mathbb{E}[\Lambda(X_a, Z_1)|Z, R = 0]
$$

$$
= \mathbb{E}[\Lambda(W)|Z, R = 0],
$$

where the second equality is by Assumption 2, and the fourth line follows by Assumption 1(e). Likewise for (4),

$$
\mathbb{E}[R1(Y \leq q_\tau)|Z] = \mathbb{E}[1(Y \leq q_\tau)|Z, R = 1] \cdot \mathbb{P}[R = 1|Z]
$$

$$
= \mathbb{E}[\Lambda(W)|Z, R = 0] \cdot r(Z)
$$

$$
= \mathbb{E}[(1 - R)\Lambda(W)|Z] \cdot \frac{r(Z)}{1 - r(Z)}.
$$

Thus, Lemma 1 follows immediately from (3) (or (4)) and Assumption 3.   ∎

Lemma A.1 *Suppose (i) $\Lambda(w; \beta)$ is measurable with respect to w for all $\beta \in \Theta_\beta$; (ii) W is bounded complete for Z, relative to the auxiliary population; (iii) $\Lambda(w; \beta)$ is differentiable with respect to β; and (iv) $\partial\Lambda(\cdot; \beta)/\partial\beta$ is uniformly bounded and $\partial\Lambda(\cdot; \beta)/\partial\beta \not\equiv 0$ for all $\beta \in \Theta_\beta$. Then, under Assumptions 1 and 2, β_0 can be uniquely identified from (3) or (4).*

Proof of Lemma A.1:   From Lemma 1, we know that β_0 solves (3) or (4). It remains to show uniqueness. Suppose, there is β_1, $\beta_1 \neq \beta_0$, that solves (3), then, $\mathbb{E}[\Lambda(W; \beta_1) - \Lambda(W; \beta_0)|Z, R = 0] = 0$. By MVT, this and (iii) implies that $\mathbb{E}[\partial\Lambda(W; \beta)/\partial\beta|_{\beta=\widetilde{\beta}}|Z, R = 0](\beta_1 - \beta_0) = 0$, for some value between β_0 and β_1. Condition (i), (ii), and (iv) then imply that $\mathbb{E}[\partial\Lambda(W; \beta)/\partial\beta|_{\beta=\widetilde{\beta}}|Z, R = 0] \not\equiv 0$, which leads to a contradiction.   ∎

Proof of Theorem 2:   We shall first prove the identification result for a fixed counterfactual distribution. Next, we take the derivative of the counterfactual experiments with respect to

t. The result of Theorem 2 then follows by the fact that Hadamard derivative operator of the quantile functional is linear. For any $t \leq t_0$, fix $\epsilon_t \in \Phi^*$, and we have that

$$
\begin{aligned}
F_{\widetilde{Y}_{s,t}|R}&(q_\tau|1) \\
&= \int P(g_s(\widetilde{X}_{s,t}, \widetilde{Z}_{1,t}, \widetilde{\epsilon}_{s,t}) \leq q_\tau | \widetilde{X}_{s,t} = x, \widetilde{Z}_{1,t} = z_1, \widetilde{R}_t = 1) dF_{\widetilde{X}_{s,t}\widetilde{Z}_{1,t}|\widetilde{R}_t}(x, z_1|1) \\
&= \int P(g_s(G_t^{-1}(\widetilde{U}_{s,t}), \widetilde{Z}_{1,t}, \widetilde{\epsilon}_{s,t}) \leq q_\tau | \widetilde{U}_{s,t} = u, \widetilde{Z}_{1,t} = z_1, \widetilde{R}_t = 1) dF_{\widetilde{U}_{s,t}\widetilde{Z}_{1,t}|\widetilde{R}_t}(u, z_1|1) \\
&= \int P(g_s(G_t^{-1}(u), Z_1, \epsilon_s) \leq q_\tau | Z_1 = z_1, R = 1) dF_{U_s Z_1|R}(u, z_1|1) \\
&= \int P(g_s(G_t^{-1}(u), Z_1, \epsilon_s) \leq q_\tau | U_s = u, Z_1 = z_1, R = 1) dF_{U_s Z_1|R}(u, z_1|1) \\
&= \int P(g_s(X_s, Z_1, \epsilon_s) \leq q_\tau | X_s = G_t^{-1}(u), Z_1 = z_1, R = 1) dF_{U_s Z_1|R}(u, z_1|1) \\
&= \int P(g_s(X_s, Z_1, \epsilon_s) \leq q_\tau | X_s = G_t^{-1}(u), Z_1 = z_1, R = 1) dF_{U_s Z|R}(u, z|1) \\
&= \int P(g_s(X_s, Z_1, \epsilon_s) \leq q_\tau | X_s = G_t^{-1}(F_{X_s}(x)), Z_1 = z_1, R = 1) dF_{X_s Z|R}(x, z|1) \\
&= \int F_{Y_s|X_s Z_1 R}(q_\tau | G_t^{-1}(F_{X_s|R}(x|1)), Z_1, 1) dF_{X_s Z|R}(x, z|1) \\
&= \int F_{Y_s|X_s Z_1 R}(q_\tau | G_t^{-1}(F_{X_s|R}(x|1)), Z_1, 1) \frac{r(z)(1-Q_0)}{Q_0(1-r(z))} dF_{XZ|R}(x, z|0) \\
&= \mathbb{E}\left[F_{Y_s|X_s Z_1 R}(q_\tau | G_t^{-1}(F_{X_s|R}(X|1)), Z_1, 1) \frac{r(Z)(1-Q_0)}{Q_0(1-r(Z))} \Big| R = 0 \right] \\
&= \frac{1}{1-Q_0} \mathbb{E}\left[(1-R)\ell(Z) \cdot F_{Y_s|X_s Z_1 R}(q_\tau | G_t^{-1}(F_{X|R}(X|1)), Z_1, 1) \right],
\end{aligned}
$$

where the second line follows by the definition of $F_{\widetilde{Y}_{s,t}}(q_\tau)$, the third one comes from the definition of $\widetilde{U}_{s,t}$ and a change of variable from x to u, the fourth equality follows by the construction of Φ^* and Assumptions 4(a) and (b), the fifth line is again by Assumption 4(a), the eighth one follows by the definition of U_s and standard change-of-variable argument, the tenth line is by Assumptions 1(a)–(c) and Bayes' Law.

To obtain the marginal distributional effect, we take derivative of $F_{\widetilde{Y}_s|R}^{G_t}(q_\tau|1)$ with respect to t and evaluate it at $t = 0$. For the marginal distributional shift,

$$
\begin{aligned}
\frac{\partial F_{\widetilde{Y}_{s,t}|R}(q_\tau|1)}{\partial t}\Big|_{t=0} &= \int \frac{\partial F_{Y_s|X_s Z_1 R}(q_\tau|x, z_1, 1)}{\partial x} \cdot \frac{\partial G_{t,p}^{-1}(F_{X_s|R}(x|1))}{\partial t}\Big|_{t=0} \\
&\qquad\qquad\qquad \cdot \frac{r(z)(1-Q_0)}{Q_0(1-r(z))} dF_{W|R}(w|0) \\
&= \frac{1}{1-Q_0} \mathbb{E}\left[(1-R)\ell(Z) \cdot \frac{\partial \Lambda(X, Z_1)}{\partial x} \cdot \frac{\partial G_{t,p}^{-1}(F_{X_s|R}(x|1))}{\partial t}\Big|_{t=0} \right].
\end{aligned}
$$

Observe that $\frac{\partial G_{t,p}^{-1}(\cdot)}{\partial t}\big|_{t=0}$ is the pathwise derivative of the inverse map $H \mapsto H^{-1}$ at $F_{X_s|R=1}$ in the direction of $G - F_{X_s|R=1}$. By Lemma 3.9.23 in Van Der Vaart and Wellner (1996), the

inverse map is Hadamard differentiable under the conditions specified in the theorem, with the derivative map given by,

$$\phi \mapsto -(\phi/h) \circ H^{-1},$$

where h is the first-order derivative of H. Let $\phi(\cdot) = G(\cdot) - F_{X_s|R=1}(\cdot)$ and $H = F_{X_s|R=1}$, it follows immediately that, for all $u \in [0,1]$,

$$\left.\frac{\partial G_{t,p}^{-1}(u)}{\partial t}\right|_{t=0} = -\frac{G(F_{X_s|R=1}^{-1}(u)) - u}{f_{X_s|R=1}(F_{X_s|R=1}^{-1}(u))},$$

and hence, for all $x \in \mathcal{X}$,

$$\left.\frac{\partial G_{t,p}^{-1}(F_{X_s|R}(x|1))}{\partial t}\right|_{t=0} = \frac{F_{X_s|R}(x|1) - G(x)}{f_{X_s|R=1}(x)}.$$

Analogously, for marginal quantile shift,

$$\begin{aligned}
\left.\frac{\partial F_{\widetilde{Y}_{s,t}|R}(q_\tau|1)}{\partial t}\right|_{t=0} &= \int \frac{\partial F_{Y_s|X_s Z_1 R}(q_\tau|x,z_1,1)}{\partial x} \cdot \left.\frac{\partial G_{t,q}^{-1}(F_{X_s|R}(x|1))}{\partial t}\right|_{t=0} \\
&\qquad\qquad \cdot \frac{r(z)(1-Q_0)}{Q_0(1-r(z))} dF_{W|R}(w|0) \\
&= \frac{1}{1-Q_0}\mathbb{E}\left[(1-R)\ell(Z) \cdot \frac{\partial \Lambda(X,Z_1)}{\partial x} \cdot (G^{-1}(F_{X|R=1}(X)) - X)\right],
\end{aligned}$$

where the second equality follows from Lemma 1 and the definition of $G_{t,q}^{-1}(\cdot)$.

To identify $F_{X_s|R=1}$, we exploit the following fact

$$\begin{aligned}
F_{X_s|R=1}(\cdot) &= \mathbb{E}[1(X \le \cdot)|R=1] \\
&= \int_{\mathcal{Z}}\int_{\mathcal{X}} 1(x \le \cdot) dF_{X|ZR}(x|z,1) dF_{Z|R}(z|1) \\
&= \int_{\mathcal{Z}}\int_{\mathcal{X}} 1(x \le \cdot) \cdot \frac{(1-Q_0)r(z)}{Q_0(1-r(z))} dF_{X|ZR}(x|z,0) dF_{Z|R}(z|0) \\
&= \mathbb{E}\left[\frac{(1-Q_0)r(Z)}{Q_0(1-r(Z))}1(X \le \cdot)|R=0\right] \\
&= \frac{1}{1-Q_0}\mathbb{E}\left[(1-R)\ell(Z)1(X \le \cdot)\right],
\end{aligned}$$

where the third line is due to Assumption 1 and Bayes' Law.

Theorem 2 then follows from Assumption 5, $q_\tau = F_{Y_s|R=1}^{-1}(\tau)$, and the fact that the Hadamard derivative of the quantile functional is $\nu_\tau'(\phi) = -\frac{\phi}{f_{Y_s|R=1}} \circ F_{Y_s|R=1}^{-1}(\tau)$, which is linear in ϕ. ∎

Proof of Theorem 3: First, we fix $U_s \in \mathcal{U}_s$ and $\phi_t \in \Phi^*$, for $t \le t_0$. By construction, there exists $\widetilde{U}_{s,t} \in \widetilde{\mathcal{U}}_{s,t}$ such that $(\widetilde{U}_{s,t}|Z_1, R=1) \overset{d}{=} (U_s|Z_1, R=1)$. Now we rewrite $F_{\widetilde{Y}_{s,t}|R=1} \in \mathcal{F}_{\widetilde{Y}_{s,t}|R=1}$ in terms of U_s and Z_1. Let $x^0 = -\infty$, and we have that

$$F_{\widetilde{Y}_{s,t}|R=1}(q_\tau)$$

$$= \int P(g_s(\widetilde{X}_{s,t}, \widetilde{Z}_{1,t}, \widetilde{\epsilon}_{s,t}) \leq q_\tau | \widetilde{X}_t = x, \widetilde{Z}_{1,t} = z_1, \widetilde{R}_t = 1) dF_{\widetilde{X}_{s,t} \widetilde{Z}_{1,t} | \widetilde{R}_t}(x, z_1 | 1)$$

$$= \sum_{j=1}^{l} \int P(g_s(\widetilde{X}_{s,t}, \widetilde{Z}_{1,t}, \widetilde{\epsilon}_{s,t}) \leq q_\tau | \widetilde{X}_{s,t} = x^j, \widetilde{Z}_{1,t} = z_1, \widetilde{R}_t = 1)$$

$$\cdot P(\widetilde{U}_{s,t} \in (G_{t,p}(x^{j-1}), G_{t,p}(x^j)] | \widetilde{Z}_{1,t} = z_1, \widetilde{R}_t = 1) dF_{\widetilde{Z}_{1,t} | \widetilde{R}_t}(z_1 | 1)$$

$$= \sum_{j=1}^{l} \int P(g_s(X_s, Z_1, \epsilon_s) \leq q_\tau | X = x^j, Z_1 = z_1, R = 1)$$

$$\cdot P(U_s \in (G_{t,p}(x^{j-1}), G_{t,p}(x^j)] | Z_1 = z_1, R = 1) dF_{Z_1 | R}(z_1 | 1)$$

$$= \sum_{j=1}^{l} \int F_{Y_s | X_s Z_1 R}(q_\tau | x^j, z_1, 1) \cdot (P(U_s \in (F_{X_s | R}(x^{j-1} | 1), F_{X_s | R}(x^j | 1)] | Z_1 = z_1, R = 1)$$

$$+ P(U_s \in (G_{t,p}(x^{j-1}), G_{t,p}(x^j)] | Z_1 = z_1, R = 1)$$

$$- P(U_s \in (F_{X_s | R}(x^{j-1} | 1), F_{X_s | R}(x^j | 1)] | Z_1 = z_1, R = 1)) dF_{Z_1 | R}(z_1 | 1)$$

$$= F_{Y_s | R}(q_\tau | 1) - \sum_{j=1}^{l} \int \Lambda(x^j, z_1)(P(U_s \in (G_{t,p}(x^{j-1}), G_{t,p}(x^j)] | Z_1 = z_1, R = 1)$$

$$- P(U_s \in (F_{X_s | R}(x^{j-1} | 1), F_{X_s | R}(x^j | 1)] | Z_1 = z_1, R = 1)) dF_{Z_1 | R}(z_1 | 1). \tag{15}$$

For the second term on the right hand side of the last equality, we have

$$P(U_s \in (G_{t,p}(x^{j-1}), G_{t,p}(x^j)] \mid Z_1 = z_1, R = 1)$$

$$- P(U_s \in (F_{X_s | R}(x^{j-1} | 1), F_{X_s | R}(x^j | 1)] \mid Z_1 = z_1, R = 1))$$

$$= (F_{U_s | Z_1 R}(G_{t,p}(x^j) \mid z_1, 1) - F_{U_s | Z_1 R}(F_{X_s | R}(x^j | 1) \mid z_1, 1)$$

$$- (F_{U_s | Z_1 R}(G_{t,p}(x^{j-1}) \mid z_1, 1)) - F_{U_s | Z_1 R}(F_{X_s | R}(x^{j-1} | 1) \mid z_1, 1))$$

$$= (G_{t,p}(x^j) - F_{X_s | R}(x^j | 1)) \cdot f_{U_s | Z_1 R}(\widetilde{u}_{j,t} \mid z_1, 1)$$

$$- (G_{t,p}(x^{j-1}) - F_{X_s | R}(x^{j-1} | 1)) \cdot f_{U_s | Z_1 R}(\widetilde{u}_{j-1,t} \mid z_1, 1)$$

$$= t \cdot (G(x^j) - F_{X_s | R}(x^j | 1)) \cdot f_{U_s | Z_1 R}(\widetilde{u}_{j,t} \mid z_1, 1)$$

$$- t \cdot (G(x^{j-1}) - F_{X_s | R}(x^{j-1} | 1)) \cdot f_{U_s | Z_1 R}(\widetilde{u}_{j-1,t} \mid z_1, 1),$$

where $\widetilde{u}_{j,t}$ is some value between $G_{t,p}(x^j)$ and $F_{X_s | R}(x^j | 1)$, and is potentially dependent on z_1.

The last equality is due to MVT. Using the above result, (15) becomes

$$F_{Y_s | R}(q_\tau | 1) - \sum_{j=1}^{l} t \cdot \int \Lambda(x^j, z_1) \cdot ((G(x^j) - F_{X_s | R}(x^j | 1)) \cdot f_{U_s | Z_1 R}(\widetilde{u}_{j,t} \mid z_1, 1)$$

$$- (G(x^{j-1}) - F_{X_s | R}(x^{j-1} | 1)) \cdot f_{U_s | Z_1 R}(\widetilde{u}_{j-1,t} \mid z_1, 1)) dF_{Z_1 | R}(z_1 | 1)$$

$$= F_{Y_s | R}(q_\tau | 1) - \sum_{j=2}^{l} t \cdot \int (\Lambda(x^{j-1}, z_1) - \Lambda(x^j, z_1))$$

$$\cdot (G(x^{j-1}) - F_{X_s|R}(x^{j-1}|1)) \cdot f_{U_s|Z_1 R}(\widetilde{u}_{j-1,t} \mid z_1, 1) dF_{Z_1|R}(z_1|1),$$

where the equality follows by rearranging terms, the fact that $G(x^0) = F_{X_s|R}(x^0|1) = 0$, and that $G(x^l) = F_{X_s|R}(x^l|1) = 1$. The pathwise derivative can thus be calculated as

$$\lim_{t \downarrow 0} \frac{F_{\widetilde{Y}_{s,t}|R=1}(q_\tau) - F_{Y_s|R}(q_\tau|1)}{t}$$

$$= \sum_{j=2}^l \int (\Lambda(x^{j-1}, z_1) - \Lambda(x^j, z_1))$$

$$\cdot (G(x^{j-1}) - F_{X_s|R}(x^{j-1}|1)) dF_{Z_1|U_s R}(z_1 \mid F_{X_s|R}(x^{j-1}|1), 1),$$

where the second line is due to the dominated convergence theorem, Bayes's Law, and the fact that $U_s|R = 1$ follows the standard uniform distribution. Therefore, by Lemma 1 and the linearity of $\nu'_\tau(\cdot)$,

$$UQE_p(\tau, G) \in \left[\inf_{U_s \in \mathcal{U}_s} \sum_{j=2}^l \int h_{q_\tau}(x^j, x^{j-1}, z_1) dF_{Z_1|U_s R}(z_1 \mid F_{X_s|R}(x^{j-1}|1), 1), \right.$$

$$\left. \sup_{U_s \in \mathcal{U}_s} \sum_{j=2}^l \int h_{q_\tau}(x^j, x^{j-1}, z_1) dF_{Z_1|U_s R}(z_1 \mid F_{X_s|R}(x^{j-1}|1), 1) \right].$$

Using a similar argument as in the proof of Theorem 5 in Rothe (2012), we can show that for $j = 1, \ldots, l$, $\{F_{Z_1|U_s R}(z_1|U_s = F_{X_s|R}(x^j|1), R = 1) : U_s \in \mathcal{U}_s\}$ is the set of all multivariate distribution functions with support equal to $\text{Supp}(F_{Z_1|R=1})$. To see this, note that for $j = 1, \ldots, l$, $F_{Z_1|U_s R}(\cdot|U_s = F_{X_s|R}(x^j|1), R = 1) = C_1^{U_s}(F_{X_s|R}(x^j|1), F_{Z_1|R}(\cdot|1))$, where the conditional copula, C^{U_s}, is defined by $C^{U_s}(F_{U_s|R}(u|1), F_{Z_1|R}(z_1|1)) := F_{U_s Z_1|R}(u, z_1|1)$, and $C_1^{U_s}$ is the partial derivative of C^{U_s} with respect to the first argument. By the construction of Φ, the set of $C^{U_s}(\cdot, \cdot)$ for $U_s \in \mathcal{U}_s$ is equivalent to the identified set of the conditional copula of X_s and Z_1 given $R = 1$, $C^{X_s}(\cdot, \cdot)$, where $C^{X_s}(F_{X_s|R}(x|1), F_{Z_1|R}(z_1|1)) := F_{X_s Z_1|R}(x, z|1)$, for all $x \in \{x^1, \ldots, x^l\}$. Then, the desired result follows by applying an extension of Theorem 2.2.7 in Nelsen (2007).

Without loss of generality, we focus on the upper bound for now. By appropriately choosing Dirac measures with unit masses on $\{z_j^*\}_{j \in \mathcal{J}_+}$ and $\{z_j^\dagger\}_{j \in \mathcal{J}_-}$, It is straightforward to show that,

$$\sup_{U_s \in \mathcal{U}_s} \sum_{j=2}^l \int h_{q_\tau}(x^j, x^{j-1}, z_1) dF_{Z_1|U_s R}(z_1 \mid F_{X_s|R}(x^{j-1}|1), 1)$$

$$= \sum_{j \in \mathcal{J}_+} h_{q_\tau}(x^j, x^{j-1}, z_{1,j}^*) + \sum_{j \in \mathcal{J}_-} h_{q_\tau}(x^j, x^{j-1}, z_{1,j}^\dagger). \tag{16}$$

The right hand side of (16) is identified under the support condition in Assumption 1(a). The

proof for the lower bound follows by an analogous argument. ■

B   Asymptotic Linear Representation of UQE Estimators

We specify additional regularity conditions in Theorem 5 and provide linear expansions for $\widehat{UQE}_p$ and $\widehat{UQE}_q$ in this section, the proofs of which are contained in the Online Appendix.

Assumption B.1

(a) $f_{X|ZR=0}$ is uniformly bounded, twice continuously differentiable with uniformly bounded first and second order derivatives on $\mathcal{X}\,\mathcal{Z}$.

(b) (i) $K_x(\cdot)$ is a second order symmetric kernel function; (ii) the support of K_x is continuous, bounded, with compact support, $K_x(\cdot)$, and such that $\int K_x(x)dx = 1$, $\int x K_x(x)dx = 0$, $\int x^2 K_x(x)dx > 0$, and $\int K_x^2(x)dx < \infty$.

(c) (i) $nb_x/log(n) \to \infty$ and (ii) $nb_x^4 \to 0$.

For $j = p, q$, the asymptotic linear representation of $\widehat{UQE}_j$ is given as follows,

$$
\widehat{UQE}_j(\tau, G) - UQE_j(\tau, G) - B_j(\tau, d, b_y)
$$
$$
= \frac{1}{n} \sum_{i=1}^{n} \left\{ \psi_{f_y,j}(A_i; \theta_0, q_\tau, G) - \frac{1}{f_{Y|R}(q_\tau|1)} \psi_{d,j}(A_i; \theta_0, q_\tau, G) \right\} + o_p(n^{-1/2} b_y^{-1/2} + b_y^2),
$$
$$
= \frac{1}{n} \sum_{i=1}^{n} \psi_j(A_i; \theta_0, q_\tau, b_y) + o_p(n^{-1/2} b_y^{-1/2} + b_y^2), \tag{17}
$$

where

$$
\psi_{f_y,j}(a; \theta_0, q_\tau, G) := \frac{d_j(\theta_0, G)}{f_{Y|R}^2(q_\tau|1)} \frac{r}{Q_0} \left(K_{b_y}(y - q_\tau) \right.
$$
$$
\left. - \mathbb{E}[K_{b_y}(Y - q_\tau)|R = 1] - \frac{(1(y \le q_\tau) - \tau) f'_{Y|R}(q_\tau|1)}{f_{Y|R}(q_\tau|1)} \right), \tag{18}
$$

$$
\psi_{d,j}(a; \theta_0, q_\tau, G) := \left(M_{\theta,j}(\theta_0)' \psi_\theta(a; \theta_0, q_\tau) \right.
$$
$$
\left. + \psi_{g,j}(a; \theta_0, G) + \frac{(1 - r)\ell(z)\Lambda_x(w; \beta_0)g_j(x)}{1 - Q_0} - \frac{r d_j(\theta_0, G)}{Q_0} \right). \tag{19}
$$

In the above equation, $\psi_\theta(a; \theta_0, q_\tau)$ are defined in the Supplementary Appendix,

$$
M_{\theta,j}(\theta_0) := \mathbb{E} \left[\frac{1 - R}{Q_0} \left(\begin{array}{c} \Lambda_x(W; \beta_0) \left(\nabla_{L,\theta_0}(Z) \cdot g_j(X) + G_{j,\theta_0}(X) \right) \\ \frac{L_0(Z)}{1 - L_0(Z)} \Lambda_{x,\beta}(W; \beta_0) g_j(X) \end{array} \right) \right], \tag{20}
$$

$$\nabla_{L,\theta}(z) := \begin{pmatrix} \dfrac{L'_s(z)(1 - L_a(z)) + L_s(z)L'_a(z)}{(1 - L_a(z))^2} \cdot k(z) \\ \dfrac{L'_s(Z)}{1 - L_a(Z)} \cdot t(Z) \\ \dfrac{L_s(Z)L'_a(Z)}{(1 - L_a(Z))^2} \cdot t(Z) \end{pmatrix}, \tag{21}$$

for $L_j(z) := L(k(z)'\gamma + t(z)'\lambda_j)$, and $L'_j(z) := L'(k(z)'\gamma + t(z)'\lambda_j)$, $j = s, a$. In addition,

$$G_{q,\theta_0}(x) := G'(G^{-1}(F_{X|R}(x|1)))^{-1} \cdot \mathbb{E}\left[\frac{1 - R}{Q_0} \cdot \nabla_{L,\theta_0}(Z)1(X \leq x)\right], \tag{22}$$

$$G_{p,\theta_0}(x) := \mathbb{E}\left[\frac{1 - R}{Q_0}\nabla_{L,\theta_0}(Z)1(X \leq x)\right] \bigg/ f_{X|R}(x|1)$$
$$+ (G(x) - F_{X|R=1}(x))\,\mathbb{E}\left[\frac{1 - R}{Q_0}\nabla_{L,\theta_0}(Z)I_{b_x}K_{b_x}(X - x)\right] \bigg/ f_{X|R}(x|1)^2,$$

$$\psi_{g,q}(a; \theta_0, G) := \mathbb{E}\left[\frac{1 - R}{1 - Q_0} \cdot \frac{\ell(Z)\Lambda_x(W; \beta_0)}{G'\left(G^{-1}(F_{X|R}(X|1))\right)}\right.$$
$$\left. \cdot \left(\frac{(1 - r)\ell(z)1(x \leq X)}{1 - Q_0} - \frac{rF_{X|R}(X|1)}{Q_0}\right)\right], \tag{23}$$

$$\psi_{g,p}(a; \theta_0, G) := \mathbb{E}\left[\frac{(1 - R)\ell(Z)\Lambda_x(W; \beta_0)}{(1 - Q_0)f_{X|R}(X|1)} \cdot \left(\frac{(1 - r)\ell(z)1(x \leq X)}{1 - Q_0} - F_{X|R}(X|1)\right)\right]$$
$$+ \frac{(1 - r)\ell(z)}{1 - Q_0}\pi(x) - \mathbb{E}\left[\frac{(1 - R)\ell(Z)}{1 - Q_0}\pi(X)\right]$$
$$- \frac{r - Q_0}{Q_0} \cdot \mathbb{E}\left[\frac{(1 - R)\ell(Z)\Lambda_x(W; \beta_0)}{1 - Q_0} \cdot \frac{G(X)}{f_{X|R}(X|1)}\right], \tag{24}$$

$$\pi(x) := \mathbb{E}\left[\Lambda_x(W; \beta_0)|\, X = x, R = 1\right]\frac{G(x) - F_{X|R}(x|1)}{f_{X|R}(x|1)}. \tag{25}$$

References

Angrist, J. D., and Krueger, A. B. (1992), "The Effect of Age at School Entry on Educational Attainment: an Application of Instrumental Variables with Moments from Two Samples," *Journal of the American Statistical Association*, 87(418), 328–336.

Arellano, M., and Meghir, C. (1992), "Female Labour Supply and On-the-Job Search: an Empirical Model Estimated Using Complementary Data Sets," *The Review of Economic Studies*, 59(3), 537–559.

Blundell, R., Chen, X., and Kristensen, D. (2007), "Semi-Nonparametric IV estimation of shape-invariant Engel curves," *Econometrica*, 75(6), 1613–1669.

Buchinsky, M., Li, F., and Liao, Z. (2022), "Estimation and inference of semiparametric models using data from several sources," *Journal of Econometrics*, 226(1), 80–103.

Canay, I. A., Santos, A., and Shaikh, A. M. (2013), "On the Testability of Identification in Some Nonparametric Models with Endogeneity," *Econometrica*, 81(6), 2535–2559.

Cattaneo, M. D., Jansson, M., and Ma, X. (2020), "Simple Local Polynomial Density Estimators," *Journal of the American Statistical Association*, 115(531), 1449–1455.

Chen, X., Hong, H., Tarozzi, A. et al. (2008), "Semiparametric Efficiency in GMM Models with Auxiliary Data," *The Annals of Statistics*, 36(2), 808–843.

Chernozhukov, V., Fernández-Val, I., and Melly, B. (2013), "Inference on Counterfactual Distributions," *Econometrica*, 81(6), 2205–2268.

Chernozhukov, V., Lee, S., and Rosen, A. M. (2013), "Intersection Bounds: Estimation and Inference," *Econometrica*, 81(2), 667–737.

Cragg, J. G., and Donald, S. G. (1996), "On the Asymptotic Properties of LDU-Based Tests of the Rank of a Matrix," *Journal of the American Statistical Association*, 91(435), 1301–1309.

Fan, Y., Sherman, R., and Shum, M. (2014), "Identifying Treatment Effects under Data Combination," *Econometrica*, 82(2), 811–822.

Firpo, S., Fortin, N., and Lemieux, T. (2009*a*), "Supplement to 'Unconditional Quantile Regressions'," *Econometrica Supplemental Material*, 77.

Firpo, S., Fortin, N. M., and Lemieux, T. (2009*b*), "Unconditional Quantile Regressions," *Econometrica*, 77(3), 953–973.

Firpo, S. P., Fortin, N. M., and Lemieux, T. (2018), "Decomposing Wage Distributions Using Recentered Influence Function Regressions," *Econometrics*, 6(2), 28.

Fortin, N., Lemieux, T., and Firpo, S. (2011), "Decomposition Methods in Economics," in *Handbook of Labor Economics*, Vol. 4, Amsterdam: Elsevier, pp. 1–102.

Graham, B. S., Pinto, C. C. d. X., and Egel, D. (2016), "Efficient Estimation of Data Combination Models by the Method of Auxiliary-to-Study Tilting (AST)," *Journal of Business & Economic Statistics*, 34(2), 288–301.

Guerre, E., Perrigne, I., and Vuong, Q. (2000), "Optimal Nonparametric Estimation of First-Price Auctions," *Econometrica*, 68(3), 525–574.

Härdle, W., and Stoker, T. M. (1989), "Investigating smooth multiple regression by the method of average derivatives," *Journal of the American statistical Association*, 84(408), 986–995.

Heckman, J. J., Lochner, L. J., and Todd, P. E. (2006), "Earnings Functions, Rates of Return and Treatment Effects: The Mincer Equation and Beyond," *Handbook of the Economics of Education*, 1, 307–458.

Hirukawa, M., Murtazashvili, I., and Prokhorov, A. (2020), "Yet Another Look at the Omitted Variable Bias,". Working Paper.

Hoeffding, W. et al. (1977), "Some Incomplete and Boundedly Complete Families of Distributions," *The Annals of Statistics*, 5(2), 278–291.

Imbens, G. W., and Lancaster, T. (1994), "Combining Micro and Macro Data in Microeconometric Models," *The Review of Economic Studies*, 61(4), 655–680.

Inoue, A., and Solon, G. (2010), "Two-Sample Instrumental Variables Estimators," *The Review of Economics and Statistics*, 92(3), 557–561.

Klevmarken, A. (1982), "Missing Variables and Two-Stage Least-Squares Estimation from More Than One Data Set," in *1981 Proceedings of the American Statistical Association, Business and Economic Statistics Section*.

Koenker, R., and Bassett Jr, G. (1978), "Regression Quantiles," *Econometrica*, 46(1), 33–50.

Lehmann, E. (1986), *Testing Statistical Hypotheses (Second Ed.)*, New York: Wiley.

Lemieux, T. (2006), "The "Mincer Equation" Thirty Years after Schooling, Experience, and Earnings," in *Jacob Mincer a pioneer of modern labor economics* Springer, pp. 127–145.

Li, T., Perrigne, I., and Vuong, Q. (2002), "Structural Estimation of the Affiliated Private Value Auction Model," *RAND Journal of Economics*, 33(2), 171–193.

Martínez-Iriarte, J. (2021), "Sensitivity Analysis in Unconditional Quantile Effects,". Working Paper.

Martinez-Iriarte, J., and Sun, Y. (2020), "Identification and Estimation of Unconditional Policy Effects of an Endogenous Binary Treatment," *arXiv preprint arXiv:2010.15864*, .

Matzkin, R. L. (2003), "Nonparametric Estimation of Nonadditive Random Functions," *Econometrica*, 71(5), 1339–1375.

Matzkin, R. L. (2007), "Nonparametric Identification," in *Handbook of Econometrics*, Vol. 6, Amsterdam: Elsevier, pp. 5307–5368.

Mincer, J. A. et al. (1974), "Schooling, Experience, and Earnings," *NBER Books*, .

Nelsen, R. B. (2007), *An Introduction to Copulas*, New York: Springer Science & Business Media.

Newey, W. K., and Powell, J. L. (2003), "Instrumental Variable Estimation of Nonparametric Models," *Econometrica*, 71(5), 1565–1578.

Powell, J. L., Stock, J. H., and Stoker, T. M. (1989), "Semiparametric Estimation of Index Coefficients," *Econometrica*, 57(6), 1403–1430.

Regan, T. L., and Oaxaca, R. L. (2009), "Work Experience as a Source of Specification Error in Earnings Models: Implications for Gender Wage Decompositions," *Journal of Population Economics*, 22(2), 463–499.

Ridder, G., and Moffitt, R. (2007), "The Econometrics of Data Combination," in *Handbook of Econometrics*, Vol. 6, Amsterdam: Elsevier, pp. 5469–5547.

Robin, J.-M., and Smith, R. J. (2000), "Tests of Rank," *Econometric Theory*, 16(2), 151–175.

Robins, J. M., Rotnitzky, A., and Zhao, L. P. (1994), "Estimation of Regression Coefficients When Some Regressors Are Not Always Observed," *Journal of the American Statistical Association*, 89(427), 846–866.

Rothe, C. (2010), "Nonparametric Estimation of Distributional Policy Effects," *Journal of Econometrics*, 155(1), 56–70.

Rothe, C. (2012), "Partial Distributional Policy Effects," *Econometrica*, 80(5), 2269–2301.

Sasaki, Y., Ura, T., and Zhang, Y. (2022), "Unconditional quantile regression with high-dimensional data," *Quantitative Economics*, 13(3), 955–978.

Van Der Vaart, A. W., and Wellner, J. A. (1996), *Weak Convergence and Empirical Processes*, New York: Springer.