TSE working paper series

09-079

Research Group: *Econometrics and Statistics*

September 2, 2009

Unconditional Partial Effects of Binary Covariates

CHRISTOPH ROTHE

Ecole d'économie de Toulouse (TSE) - Manufacture des Tabacs
Aile Jean-Jacques Laffont - 21, allée de Brienne - 31000 TOULOUSE
Tél : +33 (0) 5 61 12 85 89 - Fax : +33 (0) 5 61 12 86 37 - **www.TSE-fr.eu** - contact@TSE-fr.eu

Unconditional Partial Effects of Binary Covariates

Christoph Rothe*

Toulouse School of Economics

This Version: September 2, 2009

Abstract

In this paper, we study the effect of a small *ceteris paribus* change in the marginal distribution of a binary covariate on some feature of the unconditional distribution of an outcome variable of interest. We show that the RIF regression techniques recently proposed by Firpo, Fortin, and Lemieux (2009) do not estimate this quantity. Moreover, we show that such parameters are in general only partially identified, and derive straightforward expressions for the identified set. The results are implemented in the context of an empirical application that studies the effect of union membership rates on the distribution of wages.

JEL Classification: C14, C31

Keywords: *Unconditional Partial Effect, Partial Identification, Unconditional Quantile Regression*

*Address: Toulouse School of Economics, Manufacture des Tabacs, 21 allée de Brienne, 31000 Toulouse, France. E-Mail: c.rothe@gmx.de. Website: http://webrum.uni-mannheim.de/vwl/crothe.
I would like to thank Joshua Angrist, Andrew Chesher, Sergio Firpo, Thomas Lemieux and seminar participants at the ESEM 2009 Barcelona for helpful comments. The usual disclaimer applies.

1 Introduction

Empirical researchers are often interested in evaluating the impact of a change in the distribution of explanatory variables on some features of the unconditional distribution of an outcome variable of interest, such as e.g. its quantiles. The Unconditional Partial Effects (UPEs) recently introduced by Firpo, Fortin, and Lemieux (2009) constitute a convenient way to summarize such dependencies. For a binary covariate T, the UPE is defined as the impact of a marginal increase in the probability $p = \Pr(T = 1)$ on some feature $\Gamma(F_Y)$ of the distribution F_Y of an outcome variable Y of interest, holding everything else constant. This type of parameter is potentially of great practical relevance in various economic applications. To give an example, a researcher might be interested in the direct effect of a marginal increase in the proportion of unionionized workers on the mean or some quantile of the wage distribution, holding all other characteristics of the labor force constant.

This paper provides a detailed analysis of UPEs of binary covariates in general non-separable models. We show that these parameters are generally *not point identified* if the underlying model contains at least one additional covariate that also influences the outcome distribution. A rare exception to this rule is the UPE on the mean of the outcome distribution, which is identified if there are no interactions between the binary regressor and all other covariates. For most other features of the outcome distribution usually considered in applied work, including e.g. quantiles or related statistics like interquantile ranges, UPEs are typically not point identified even for extremely simplistic models, such as linear models with i.i.d. errors and no interaction terms.

In cases where point identification fails, the UPE can often still be bounded, with the identified set taking the form of an interval. We derive explicit expressions for the upper and lower bounds of this set, and show how to estimate them nonparametrically in the context of an empirical application that analyzes the effect of unionization on the wage distribution. Our empirical results suggest that a marginal change in the proportion of unionized workers has an ambiguous effect, manifested in relatively wide bounds for the parameters of interest. In particular, we cannot rule out that such a change would leave the distribution of wages entirely unaffected.

UPEs have been recently introduced to the literature by Firpo, Fortin, and Lemieux (2009). For a continuously distributed covariate, they are defined as the effect of a marginal shift in the regressor's location on the distribution of the outcome variable. For this case, Firpo, Fortin, and Lemieux establish identification of the UPE under general conditions, and propose various estimation procedures. They also consider UPEs of a binary covariate, particularly in an earlier working paper version (Firpo, Fortin, and Lemieux 2007), but only cover the case where *no additional regressors* are present in the model explicitly.

The main contribution of our paper is to show that these results do not extend to the practically relevant case of a multidimensional covariate vector. The reason is that there exists no unique way to link a multivariate distribution function to its one-dimensional marginals if at least one of them is discrete. This follows from a result in copula theory known as Sklar's Theorem (Sklar 1959). Hence, there are several ways to implement a change in the marginal distribution of a binary covariate while "holding everything else constant", and consequently, the impact of such a change on distribution of the outcome variable is only partially identified in general.

While Firpo, Fortin, and Lemieux (2009) provide a theoretical discussion of UPEs of binary covariates only for the univariate case, their empirical application (which is the same as ours) presents point estimates of such a parameter in a setting with many additional regressors. Our paper discusses the interpretation of these results. We argue that their estimates correspond in fact to a different parameter, which can be seen as a generalization of the usual Average Partial Effect. While this parameter is point identified under general conditions, it cannot be interpreted as the impact of a marginal *ceteris paribus* change in the unconditional distribution of the binary covariate.

We view our paper as an important complement to a growing literature that analyzes the impact of counterfactual changes in the distribution of covariates on the unconditional distribution of an outcome variable of interest. Examples include DiNardo, Fortin, and Lemieux (1996), Donald, Green, and Paarsch (2000), Machado and Mata (2005), Chernozhukov, Fernandez-Val, and Melly (2009) and Rothe (2009). In these papers, the focus is on the effect of a fixed change of the entire covariate distribution on the distribution of the outcome variable. The parameters they estimate are thus quite different from the ones considered in Firpo, Fortin, and Lemieux (2009) and the present paper, which correspond to marginal changes in the distribution of a single covariate.

The plan of the paper is as follows. In the next section we describe our modelling framework and the parameters of interest. Section 3 contains the identification analysis. Section 4 discusses the interpretation of the point estimates in Firpo, Fortin, and Lemieux (2009). Section 5 contains our empirical application. Finally, Section 6 concludes. Technical arguments are delegated to the Appendix.

2 Model and Parameters of Interest

The setup we consider is as follows[1]: we observe a scalar dependent variable Y and a $(d+1)$-dimensional vector of covariates $Z = (T, X')'$, with marginal distribution functions F_Y and F_Z, respectively. The covariate vector consists of a dummy variable $T \in \{0, 1\}$, and the d-dimensional vector of remaining covariates X, which can be either continuously or discretely distributed. The corresponding marginal distribution functions are denoted by F_T and F_X, respectively. The dependent variable is assumed to be generated through the nonseparable model

$$Y = m(Z, \eta), \tag{2.1}$$

[1]Our setup is the same as in Firpo, Fortin, and Lemieux (2009), but we introduce it in a slightly different manner which is convenient for our later analysis.

where η is an unobserved error term that is assumed to be independent of Z [2]. In a typical microeconometric application, Z and η would correspond to observed and unobserved characteristics of an individual, respectively, and m would describe the decision rule that, given individual characteristics, determines the individual's choice Y. This flexible formulation allows the covariates to exert influence on Y in manifold ways. For example, model (2.1) allows for heteroskedasticity or skewness in the conditional distribution of Y given Z.

The parameters we are interested in correspond to the effect of certain infinitesimal perturbations of the covariates' distribution on some feature $\Gamma(F_Y)$ of the distribution of the outcome variable, where $\Gamma : \mathcal{F}_\Gamma \to \mathbb{R}$, and $\mathcal{F}_\Gamma$ is a class of distribution functions such that $F_Y \in \mathcal{F}_\Gamma$ if $|\Gamma(F_Y)| < \infty$. That is, suppose we have a sequence of distribution functions $F_{Z,\delta}$ indexed by $\delta \in \mathbb{R}$, such that

$$\lim_{\delta \to 0} \|F_{Z,\delta} - F_Z\|_\infty = 0.$$

Let Z_δ be a sequence of random vectors with distribution $F_{Z,\delta}$ that are independent of η, and define the counterfactual random variables Y_δ as

$$Y_\delta = m(Z_\delta, \eta).$$

The corresponding CDF of Y_δ, denoted $F_{Y,\delta}$, can then be written as

$$F_{Y,\delta}(y) = \int F_{Y|T,X}(y,t,x)dF_{Z,\delta}(t,x),$$

since η is assumed to be independent of both Z and Z_δ (see e.g. Rothe (2009)). With this notation, we can define the *effect of an infinitesimal perturbation of the covariate distribution in the direction* $F_{Z,\delta}$ *on* $\Gamma(F_Y)$ as the derivative of $\Gamma(F_{Y,\delta})$ with respect to δ

[2] Here we follow Firpo, Fortin, and Lemieux (2009) and do not consider models with endogeneity.

evaluated at $\delta = 0$,

$$\theta(\Gamma, F_{Z,\delta}) = \left.\frac{\partial \Gamma(F_{Y,\delta})}{\partial \delta}\right|_{\delta=0,}$$

provided that the latter quantity is well-defined.

In their paper, Firpo, Fortin, and Lemieux (2009) introduce the Unconditional Partial Effect (UPE) of a regressor T on $\Gamma(F_Y)$ as the effect of a specific infinitesimal perturbation of the covariate distribution. For continuously distributed T, it is the effect of a perturbation in the direction $F_{Z,\delta}(t,x) = F_Z(t-\delta, x)$, which corresponds to a location shift in T. In the case of a single binary covariate T, it is defined as the effect of a perturbation in the direction $F_{Z,\delta}(t) \equiv F_{T,\delta}(t) = \mathbb{I}\{0 \leq t < 1\}(F_T(0) - \delta) + \mathbb{I}\{t \geq 1\}$, corresponding to an increase in $p = \Pr(T = 1)$ to $p + \delta$. See Corollary 1 in the published version of their paper and Corollary 3 in Firpo, Fortin, and Lemieux (2007), an the earlier working paper version, respectively, for details. For a general multivariate setting, the UPE of a binary covariate is only informally defined as the effect of a perturbation corresponding to an increase in $p = \Pr(T = 1)$ to $p + \delta$, leaving all remaining features of the distribution of Z unchanged. One of the main problems we aim to solve in this paper is to give an explicit method to construct such a perturbation.

In order to accomplish this in a multivariate setting, we use a well-known result from the theory of copula functions due to Sklar (1959). Copula functions are popular tools in various areas of applied statistics and economics, including such diverse fields as finance, risk management or meteorology, since they allow researchers to model the dependence structure and the one-dimensional marginals of a multivariate distribution separately (see Joe (1997), Nelsen (2006) or Trivedi and Zimmer (2007) for extensive surveys of the related literature). This feature also makes them very attractive for our purposes. In particular, it follows from Sklar's Theorem that for every multivariate distribution function F_Z with marginal distribution functions $F_T, F_{X_1}, \ldots, F_{X_d}$ there exists a function

C^* such that

$$F_Z(t,x) = C^*(F_T(t), F_{X_1}(x_1), \ldots, F_{X_d}(x_d)), \tag{2.2}$$

where $C^* : [0,1]^{d+1} \to [0,1]$ is a multivariate distribution function with uniform one-dimensional marginal distributions. This so-called copula function C^* connects (or "couples") a multivariate CDF to its one-dimensional marginals, and can thus be interpreted as the object that contains all the information about the dependence structure of the random vector Z.

It is important to note that (2.2) is not merely a statistical representation, but also has an intuitive economic interpretation. Suppose that for every member of the population the difference in utility between choosing $T = 1$ over $T = 0$ is given by a continuously distributed random variable $\tilde{U}$, so that $T = \mathbb{I}\{\tilde{U} \geq 0\}$. Denoting the corresponding distribution function by $\tilde{F}_U$, we can define the rank of an individual in the distribution of latent utility as $U = \tilde{F}_U(\tilde{U})$. Now suppose for simplicity that X is continuously distributed, and let $V = (F_{X_1}(X_1), \ldots, F_{X_d}(X_d))'$ be the vector of corresponding ranks. Then the copula C^* in (2.2) is the distribution function of (U, V). That is, it determines the joint distribution of the ranks in the population.

Given the representation in (2.2), it appears natural to define a perturbed distribution of the covariates where only the probability of observing $T = 1$ has changed from p to $p + \delta$ as

$$G_\delta(t,x) = C^*(F_{T,\delta}(t), F_{X_1}(x_1), \ldots, F_{X_d}(x_d)),$$

where $F_{T,\delta}$ is the new marginal CDF of the first component, given by

$$F_{T,\delta}(t) = \mathbb{I}\{0 \leq t < 1\}(1 - p - \delta) + \mathbb{I}\{t \geq 1\}.$$

Note that for any $\delta \neq 0$ the only difference between G_δ and F_Z is the marginal distribution of the first component. The remaining marginals are the same for both distribution functions, and since they share the same copula function both distributions also exhibit the same dependence structure. With this notation, we can now define the UPE of a dummy variable T on $\Gamma(F_Y)$ as

$$\alpha(\Gamma, T) = \theta(\Gamma, G_\delta).$$

The following theorem gives conditions under which this parameter is a well-defined feature of the underlying data generating process, and derives an explicit representation.

Theorem 1. *Suppose that i) the real-valued functional* Γ *is Hadamard differentiable at* F_Y*, with derivative* Γ'*, ii) the copula function* C^* *is differentiable with respect to its first component, and iii) the support of* X *conditional on* $T = t$ *does not vary with* $t \in \{0, 1\}$*. Then the* Unconditional Partial Effect of a dummy variable T on $\Gamma(F_Y)$ *exists and can be written as*

$$\alpha(\Gamma, T) = \int g_\Gamma(x) ds^*(F_{X_1}(x_1), \ldots, F_{X_d}(x_d)), \tag{2.3}$$

where

$$g_\Gamma(x) = \Gamma'(F_{Y|T,X}(\cdot, 1, x)) - \Gamma'(F_{Y|T,X}(\cdot, 0, x)) \tag{2.4}$$

and

$$s^*(b) = \left.\frac{\partial C^*(a, b)}{\partial a}\right|_{a=F_T(0).} \tag{2.5}$$

The role of both condition i) and ii) in the preceding theorem is to ensure that $\Gamma(F_{Y,\delta})$ is differentiable with respect to δ. The Hadamard differentiability condition requires the functional of interest to be sufficiently smooth around F_Y in some appropriate sense. To be precise, this means that there exists a continuous linear functional Γ' such that

$$\left|\frac{\Gamma(F_Y + \delta h_\delta) - \Gamma(F_Y)}{\delta} - \Gamma'(h)\right| \to 0 \quad \text{as } \delta \to 0$$

for all functions $h_\delta \to h$ (see Van der Vaart (2000, p. 296) for details). This condition can be verified for most functionals that are commonly of interest in applied work under mild additional regularity conditions. Examples include moments, quantiles, interquantile ranges, the Lorenz curve, the Gini coefficient and other measures of inequality (see e.g. Rothe (2009)). We discuss the case of the mean and the quantiles in greater detail below. Finally, condition iii) in the preceeding theorem ensures that the conditional distribution function $F_{Y|Z}$, which enters the term g_Γ, is well-defined over the area of integration in (2.3).

3 Identification

Given knowledge of the copula function C^*, it would be straightforward to compute $\alpha(\Gamma, T)$ using the representation in Theorem 1. Unfortunately, while Sklar's Theorem guarantees the existence of a copula function satisfying (2.2), it does not ensure its uniqueness. Instead, as one can easily see, C^* is identified by the data only on the range of the marginal distribution functions $F_T, F_{X_1}, \ldots, F_{X_d}$. (see also Nelsen (2006, Theorem 2.3.3)). In particular, for any value $b_0 \in \text{Ran}(F_{X_1}) \times \ldots \times \text{Ran}(F_{X_d})$, the value $C^*(a, b_0)$ is uniquely determined for $a \in \{0, F_T(0), 1\}$ only. This in turn implies that the function s^* defined in (2.5) is not point identified, since the identification of a derivative at a fixed point requires knowledge of the function at least in some small neighbourhood.

Although s^* is not point identified, one can use the properties of copula functions to find restrictions on its shape. The following lemma establishes that s^* belongs to a very specific class of functions.

Lemma 1. *Suppose that $F_T(0) \in (0,1)$. Then $s^*(\cdot) \in \mathcal{S}$, where $\mathcal{S}$ is the set of all* *multivariate distribution functions with support* $R_{\mathcal{X}} = \{(F_{X_1}(x_1), \ldots, F_{X_d}(x_d))' : x \in \mathcal{X}\}$*, where* $\mathcal{X}$ *denotes the support of* X*.*

This result allows us to construct the identified set $\mathcal{A}(\Gamma, T)$, which contains the possible values of the UPE that are compatible with the distribution of observable quantities. Using the representation for $\alpha(\Gamma, T)$ given in (2.3), we obtain that

$$\alpha(\Gamma, T) \in \mathcal{A}(\Gamma, T) = \left\{ \int g_\Gamma(x) ds(F_{X_1}(x_1), \ldots, F_{X_d}(x_d)), s \in \mathcal{S} \right\}. \tag{3.1}$$

This expression for the identified set can be further simplified by noting that due to the properties of the functions contained in $\mathcal{S}$, we can interpret $\mathcal{A}(\Gamma, T)$ as the set of all weighted averages of the function g_Γ. If this function is is bounded, then every such weighted average is necessarily smaller than the smallest upper bound on $g_\Gamma(x)$, and bigger than the biggest lower bound. The next theorem formalizes this idea.

Theorem 2. *Suppose that the conditions of Theorem 1 hold. Then the identified set for* $\alpha(\Gamma, T)$*, the Unconditional Partial Effect of a dummy variable* T *on* $\Gamma(F_Y)$*, is given by*

$$\mathcal{A}(\Gamma, T) = [\alpha^L(\Gamma, T), \alpha^U(\Gamma, T)]$$

where

$$\alpha^U(\Gamma, T) = \sup_{x \in \mathcal{X}} g_\Gamma(x)$$
$$\alpha^L(\Gamma, T) = \inf_{x \in \mathcal{X}} g_\Gamma(x)$$

and $\mathcal{X}$ *denotes the support of* X*.*

The theorem shows that the identified set $\mathcal{A}(\Gamma, T)$ takes the form of an interval, and provides explicit expressions for its upper and lower bounds, which are easy to evaluate. Since the identified set is restricted by the extrema of the "bound generating function" $g_\Gamma(x)$, our problem falls into the general class of models with partially identified parameters restricted by intersection bounds. A general theory for estimation and inference in this setting is provided by Chernozhukov, Lee, and Rosen (2009), whose results we also use in our empirical application below.

It is an immediate consequence of Theorem 1 that $\alpha(\Gamma, T)$ is identified if and only if $g_\Gamma(x)$ is constant for all $x \in \mathcal{X}$. In this case the upper and lower bound coincide, and the identified set reduces to a singleton. Under other circumstances, point identification necessarily fails. However, the bounds may be informative, in the sense that $\alpha^U(\Gamma, T)$ and $\alpha^L(\Gamma, T)$ are finite, if $g_\Gamma(x)$ is bounded over $x \in \mathcal{X}$. Whether or not that is the case depends on the specific form of the conditional distribution of Y given X and T, and the functional of interest Γ.

We now discuss two examples that illustrate the application of Theorem 2: the UPE on the mean and on the τ-quantile of Y.

Example 1 (Mean)**.** Suppose that $\Gamma_M(F_Y) = \int y\, dF_Y(y)$ is the functional that maps a CDF into the corresponding mean. Since this functional is linear, it is also Hadamard differentiable with derivative $\Gamma'_M = \Gamma_M$. The bounds given in Theorem 2 can thus be written as

$$\begin{aligned}
\alpha^U(\Gamma_M, T) &= \sup_{x\in\mathcal{X}} \left(\mathbb{E}(Y|T=1, X=x) - \mathbb{E}(Y|T=0, X=x)\right) \text{ and} \\
\alpha^L(\Gamma_M, T) &= \inf_{x\in\mathcal{X}} \left(\mathbb{E}(Y|T=1, X=x) - \mathbb{E}(Y|T=0, X=x)\right).
\end{aligned}$$

This implies that $\alpha(\Gamma_M, T)$ is identified whenever the conditional expectation of Y given T and X does not contain any interaction terms between T and the other regressors, i.e. it holds that $\mathbb{E}(Y|T=t, X=x) = m_1(t) + m_2(x)$. When T exerts a heterogeneous effect varying with X point identification fails. For example, the UPE is only partially identified for the Probit model. There the conditional expectation function $\mathbb{E}(Y|T = t, X=x) = \Phi(\gamma_1 + \gamma_2 t + \gamma_3' x)$, where $\Phi(\cdot)$ is the standard normal CDF, is not additively separable in t. Such a lack of additive seperability is also present in other generalized linear models and most nonlinear regression models.

Example 2 (Quantile)**.** Suppose that $\Gamma_{Q,\tau}(F_Y) = \inf\{y \in \mathbb{R} : F_Y(y) \geq \tau\} = Q_Y(\tau)$ is the functional that maps a CDF into the corresponding τ-quantile. If F_Y is continuously differentiable in some open neighbourhood of $Q_Y(\tau)$, and the derivative f_Y is strictly positive, it follows from Lemma 21.4 in Van der Vaart (2000) that Γ is Hadamard differentiable with derivative

$$\Gamma'_{Q,\tau} : \phi \mapsto -\left(\frac{\phi}{f_Y}\right) \circ Q_Y.$$

In this case the bounds given in Theorem 2 simplify to

$$\begin{aligned}\alpha^U(\Gamma_{Q,\tau}, T) &= \sup_{x\in\mathcal{X}} -\frac{F_{Y|T,X}(Q_Y(\tau)|1,x) - F_{Y|T,X}(Q_Y(\tau)|0,x)}{f_Y(Q_Y(\tau))} \text{ and} \\ \alpha^L(\Gamma_{Q,\tau}, T) &= \inf_{x\in\mathcal{X}} -\frac{F_{Y|T,X}(Q_Y(\tau)|1,x) - F_{Y|T,X}(Q_Y(\tau)|0,x)}{f_Y(Q_Y(\tau))}.\end{aligned}$$

Inspection of the bounds reveals that the UPE of a dummy variable T on the τ-quantile of the outcome distribution is *not identified* even for very simple models *without* interaction effects. Consider for example the case that $Y = T + X + \eta$. Then the numerator in the expression for the bounds is given by $F_{Y|T,X}(Q_Y(\tau)|1,x) - F_Y(Q_Y(\tau)|0,x) = F_\eta(Q_Y(\tau) - x - 1) - F_\eta(Q_Y(\tau) - x)$, which will generally depend on x.[3] On the other hand, since every distribution function is bounded between 0 and 1, it is ensured that both $\alpha^L(\Gamma_{Q,\tau}, T)$ and $\alpha^U(\Gamma_{Q,\tau}, T)$ are finite. The bounds are thus necessarily informative, although this does not guarantee that they will be narrow in a particular application.

4 What do RIF Regressions Estimate?

As mentioned above, Firpo, Fortin, and Lemieux (2009, 2007) explicitly only discuss the definition and identification of UPEs of binary regressors for the case that there are no additional covariates present in the model. However, in their empirical application they

[3]The only exception would be the rare and arguably unrealistic case that the distribution function of η is linear over the respective range of x.

report point estimates of the effect of unionization on the mean and various quantiles of the wage distribution, while controlling for a number of further human capital variables. These estimates are obtained using several variants of their so-called RIF regression techniques. Since our analysis suggests that such parameters are generally not identified, it is useful to clarify how their results can be interpreted. We now show that although they are interpreted by Firpo, Fortin, and Lemieux as if they were estimates of the UPE, they correspond in fact to a different parameter, that coincides with our UPE if and only if the copula function C^* has the same local properties as a copula that induces independence between T and X.

Using our notation, the population quantity corresponding to the point estimates in Firpo, Fortin, and Lemieux (2009) is given by[4]:

$$\beta(\Gamma, T) = \int g_\Gamma(x) dF_X(x). \tag{4.1}$$

Note that if $\Gamma = \Gamma_M$ is the functional that maps a CDF into its mean, this parameter simplifies to $\beta(\Gamma_M, T) = \mathbb{E}[\mathbb{E}(Y|T=1,X) - \mathbb{E}(Y|T=0,X)]$, which is the usual Average Partial Effect of a binary covariate (Wooldridge 2002, Chapter 2). For a general functional Γ, we therefore refer to $\beta(\Gamma, T)$ in the following as the Generalized Average Partial Effect (GAPE).

The GAPE is conceptionally different from the UPE, and cannot be interpreted as the effect of a marginal change in the unconditional probability of observing $T = 1$. While the GAPE can be written as the effect of an infinitesimal perturbation of the covariate distribution, the direction of the perturbation differs from the one used to construct the UPE. In particular, using the notation from Section 2, we have that $\beta(\Gamma, T) = \theta(\Gamma, \tilde{G}_\delta)$, where

$$\tilde{G}_\delta(t, x) = (F_{T|X}(t, x) - \delta\mathbb{I}\{0 \le t < 1\})F_X(x).$$

[4]This is not explicitly stated in the paper, but can be inferred from the Supplemental Material.

Here $\tilde{G}_\delta(t,x)$ is a perturbed covariate distribution where the *conditional* probability of observing $T=1$ given $X=x$ is changed from $p(x)=\Pr(T=1|X=x)$ to $p(x)+\delta$ for *every* value of $x \in \mathcal{X}$. While in this case the unconditional probability of observing $T=1$ increases by δ as well, in general this perturbation does not leave the original dependence structure of the covariate distribution unaffected.

While in general the GAPE and the UPE are two different parameters, they can be shown to coincide under a very specific restriction on the copula function C^*. Recall from (2.3) that the UPE can be written as

$$\alpha(\Gamma, T) = \int g_\Gamma(x) ds^*(F_{X_1}(x_1), \ldots, F_{X_d}(x_d)),$$

where $s^*(b) = \partial_a C^*(a,b)$ evaluated at $a = F_T(0)$. Comparing this expression to the term on the righ-hand side of (4.1), we see that the GAPE and the UPE are equal if the function s^* satisfies the relationship $s^*(b) = C^*(1,b)$ for all $b \in [0,1]^d$. From the definition of s^*, it follows that for each element C of the class of copula functions which imply this relationship it holds that

$$C(a,b) = aC(1,b) + o(\|a - F_T(0)\|), \tag{4.2}$$

as $a \to F_T(0)$, uniformly over $b \in R_{\mathcal{X}}$. Every element of this class therefore locally behaves in the same way as a copula function that induces independence between T and X, i.e. that has $C(a,b) = aC(1,b)$ for all a and b. One can thus think of relationship (4.2), which implies equality of UPE and GAPE, as a local independence condition. Note that imposing this condition by assumption would be sufficient to achieve point identification of the UPE. However, such an approach would typically not be attractive in practice. First, the local independence condition is not a testable property, and second it is unlikely to be justifiable in applications by economic arguments, except if T and X are fully independent.

Finally, we remark that although the UPE and the GAPE are generally different, the GAPE is always contained in the identified set $\mathcal{A}(\Gamma, T)$, since by Lemma 1 we have that $C^*(1, \cdot) \in \mathcal{S}$, the set of feasible values of the function s^*.

5 Empirical Application

In this section, we revisit the empirical application in Firpo, Fortin, and Lemieux (2009), which investigates the direct effect of unionization on the distribution of male (log) wages. We employ the same dataset, which consists of 266,956 observations on U.S. males from the 1983—1985 Outgoing Rotation Group (ORG) supplement of the Current Population Survey. Following Firpo, Fortin, and Lemieux (2009), we use a model of wage determination of the form

$$Y = m(T, X, \eta)$$

where Y is the log wage of the individual, T is an indicator for membership in a union, and X is a vector of further control variables, which include indicators for being married and being non-white, six indicators for different levels of education, and nine indicators for different levels of labour market experience. The parameters of interest are the effects of a marginal increase in the unionization rate on the mean and the quantiles of the distribution of log wages. Our above analysis suggests that these parameters are not point identified, but can be bounded. Since the support of the covariates is finite, the bounds will be informative.

In order to estimate the bounds on the UPE, we use a methodology proposed by Chernozhukov, Lee, and Rosen (2009). They consider the general problem of conducting inference on a partially identified parameter when the bounds of the identified set are given by the extrema of estimateable functions. Since in our application the bounds on the UPE are the maximum and minimum of the bound generating function $x \mapsto g_\Gamma(x)$ over the finite set $\mathcal{X}$, they fit exactly into this framework.

Table 1: Effect of Union Status on Mean and Quantiles of Log Wage Distribution

	Mean	10th Centile	50th Centile	90th Centile
UPE				
Bounds	[-0.198, 0.469]	[-0.012,0.774]	[-0.023, 0.659]	[-1.058, 0.116]
95% CI	[-0.241, 0.545]	[-0.643, 1.029]	[-0.093, 0.976]	[-1.629, 0.390]
GAPE				
Estimate	0.179	0.197	0.341	-0.136
95% CI	[0.175, 1.183]	[0.193, 0.201]	[0.333, 0.349]	[-0.144, -0.128]

Chernozhukov, Lee, and Rosen (2009) argue that simple sample analogue estimators of the bounds can be severely biased in finite samples. They therefore propose to add a precision-correction term to a suitable estimate $x \mapsto \hat{g}(x)$ of the bound-generating function before applying the maximum and minimum operators in order to obtain median unbiased estimates. They also show that a similar idea can be used to to construct asymptotically valid confidence intervals for the true parameter of interest. Since in our application all covariates are discretely distributed with finite support, these procedures can easily be implemented in a fully nonparametric fashion by using the ordinary frequency method. The details are described in Appendix B.

We apply the estimators and inference procedures to the 1983–1985 CPS data. In Table 1, we report estimates of the identified set of the UPE of union status on the mean and the 10th, 50th, and 90th quantiles of the log wage distribution, together with the respective 95% confidence intervals for the true parameter. The results are compared with the RIF-OLS estimates from Firpo, Fortin, and Lemieux (2009). In addition to that, Figure 1 shows the estimated identified sets and 95% confidence intervals of the UPE of union status on 19 different quantiles (from the 5th to the 95th). Again, these results are compared with the RIF-OLS estimates.

Our nonparametric bounds for the UPE of unionization turn out to be quite wide for all statistics we consider. The estimate of the identified set for the mean effect is $[-0.198, 0.469]$, allowing for a wide range of possible values. The upper bound on the quantile effect is highly nonmonotonic, increasing from 0.4 at the 5th quantile to 1.5 at

Figure 1: Nonparametric bounds on UPE of union status on the quantiles of log wages (shaded area) with corresponding confidence intervals (dashed area); and estimated GAPE from Firpo, Fortin, Lemieux (2009) (solid line).

the 15th quantile, then steadily declines to about 0.1 at the 90th quantile, before sharply increasing to 0.9 at the 95th quantile. In contrast, the lower bound stays roughly constant around zero from the 5th quantile to the median, and then sharply declines to -1.6 at the 95th quantile. The confidence intervals for the true parameter include the value 0 at every quantile, and thus do not rule out that a marginal change in unionization would have no effect whatsoever on the distribution of log wages.

Based on their point estimates, Firpo, Fortin, and Lemieux (2009) come to a quite different conclusion. The unconditional quantile effect they estimate exhibits an inverse U-shape, first increasing from about 0.1 at the 5th quantile to about 0.4 at the 35th quantile, before declining and eventually reaching a negative effect of about -0.2 at the 95th quantile. Firpo, Fortin, and Lemieux (2009) interpret these results as if they were estimates of the UPE, arguing that they provide evidence that "unionization progressively increases wages in the three lower quintiles of the distribution, peaking around the 35th quantile, and actually reduces wages in the top quintile of the distribution" (p. 966).

However, as described in Section 4, the GAPE parameter they actually estimate is generally different from the UPE, and does not warrant such an interpretation. More precisely, in the present context the GAPE corresponds to the effect of a small increase in unionization by *exactly the same amount* in every subgroup of the population defined by the covariates X. It would thus coincide with the UPE only if union membership rates generally changed by the same absolute amount in e.g. all educational groups or age groups. Since such uniform changes in unionization patterns have not been observed in the US or other industrialized countries in the past, this is unlikely to be a realistic assumption. The GAPE thus cannot be used to establish a direct link between unconditional union membership rates and the distribution of wages.

Instead, our interval estimates of the UPE show that the direct role of unionization is much more ambiguous, and do not rule out the possibility that changes in overall union membership rates could leave the aggregate wage distribution entirely unaffected. The reason for this ambiguity is that covariates other than union membership play a substantial role in the determination of wages. In the presence of such individual heterogeneity in the population, the effect of say a decline in unionization critically depends on *which individuals are actually leaving the unions.* In our framework, the component responsible for this relationship is the copula C^*, which governs the dependence structure between union membership and all other characteristics. Since this function is not fully identified by cross-sectional data, one cannot determine exactly how a change in the overall unionization rate would affect the unionization rate in every subgroup of the population defined through their value of the other explanatory covariates. Thus its impact on the unconditional wage distribution is only partially identified.

6 Conclusions

In this paper, we study the effect of an infinitesimal change in the marginal distribution of a binary covariate on some feature of the unconditional distribution of an outcome variable of interest, holding everything else constant. We show that such parameters are only partially identified in general, and provide an explicit expression for the identified set. We implement these results in the context of an empirical application that studies the effect of unionization on the distribution of wages.

A Proofs

Proof of Theorem 1. Our proof consists of three steps. First, it follows from the differentiability of the copula that

$$\begin{aligned}\lim_{\delta\to 0}\frac{F_{Z,\delta}(t,x)-F_Z(t,x)}{\delta} &= \mathbb{I}\{0\le t<1\}\lim_{\delta\to 0}\delta^{-1}[C^*(F_T(0)-\delta,F_{X_1}(x_1),\ldots,F_{X_d}(x_d))\\ &\quad - C^*(F_T(0),F_{X_1}(x_1),\ldots,F_{X_d}(x_d))]\\ &= -\mathbb{I}\{0\le t<1\}s^*(F_{X_1}(x_1),\ldots,F_{X_d}(x_d)).\end{aligned}$$

Second, using the previous result and the continuous mapping theorem, we obtain that

$$\begin{aligned}\lim_{\delta\to 0}\frac{F_{Y,\delta}(y)-F_Y(y)}{\delta} &= \lim_{\delta\to 0}\frac{\int F_{Y|T,X}(y,t,x)dF_{Z,\delta}(t,x)-\int F_{Y|T,X}(y,t,x)dF_Z(t,x)}{\delta}\\ &= \int F_{Y|T,X}(y,t,x)d\left(\lim_{\delta\to 0}\frac{F_{Z,\delta}(t,x)-F_Z(t,x)}{\delta}\right)\\ &= -\int F_{Y|T,X}(y,t,x)d\left(s^*(F_{X_1}(x_1),\ldots,F_{X_d}(x_d))\mathbb{I}\{0\le t<1\}\right)\\ &= \int F_{Y|T,X}(y,1,x)-F_{Y|T,X}(y,0,x)ds^*(F_{X_1}(x_1),\ldots,F_{X_d}(x_d)).\end{aligned}$$

Finally, Hadamard differentiability of Γ implies that

$$\begin{aligned}\lim_{\delta\to 0}\frac{\Gamma(F_{Y,\delta})-\Gamma(F_Y)}{\delta} &= \Gamma'\left(\int F_{Y|T,X}(\cdot,1,x)-F_{Y|T,X}(\cdot,0,x)ds^*(F_{X_1}(x_1),\ldots,F_{X_d}(x_d))\right)\\ &= \int \Gamma'(F_{Y|T,X}(\cdot,1,x))-\Gamma'(F_{Y|T,X}(\cdot,0,x))ds^*(F_{X_1}(x_1),\ldots,F_{X_d}(x_d)),\end{aligned}$$

where the last equality follows from the linearity of Γ'. □

Proof of Lemma 1. This follows from Theorem 2.2.7 in Nelsen (2006), by straightforward extension of the arguments given there from the bivariate to the general multivariate case. □

Proof of Theorem 2. Let $\mathcal{H} = \{H : H(x) = s(F_{X_1}(x_1),\ldots,F_{X_d}(x_d)), s\in\mathcal{S}\}$. Note that it follows from the properties of $\mathcal{S}$ that $\mathcal{H}$ is the set of all distribution functions with support $\mathcal{X}$. It then follows directly that

$$\inf_{x\in\mathcal{X}} g_\Gamma(x) \leq \sup_{H\in\mathcal{H}}\int g_\Gamma(x)dH(x) \leq \sup_{x\in\mathcal{X}} g_\Gamma(x).$$

Since $\mathcal{H}$ is the set of *all* distribution functions with support $\mathcal{X}$, these bounds are sharp. □

B Estimation and Inference

In this section, we describe how to construct median unbiased estimates of the bounds on the UPE, and how to obtain asymptotically valid confidence intervals for the parameter of interest. We heavily rely upon recent results by Chernozhukov, Lee, and Rosen (2009) - henceforth CLR - who provide a general theory for estimation and inference in models with partially identified parameters restricted by intersection bounds. This class includes our setting as a special case. We first explain the general principles, and then consider the cases of the mean and quantile UPE in greater detail.

B.1 General Principles

The basic idea of CLR is to add suitable precision-correction terms to a standard estimate of the bound generating function g_Γ before applying the maximum or minimum operator. To explain this in detail, we first have to introduce some notation.[5] For any $p\in(0,1)$,

[5]Note that our notation slightly differs from the one in CLR since in their paper the upper bound of the identified set is given by the infimum of the bound generating function, whereas in our case it is given by its supremum. One could simply transfer our notation back into theirs by considering the negative version of the bound generating function

we define

$$\hat{\alpha}_p^U = \max_{x \in \hat{\mathcal{X}}^U} [\hat{g}(x) - k_p s(x)] \quad \text{and} \quad \hat{\alpha}_p^L = \min_{x \in \hat{\mathcal{X}}^L} [\hat{g}(x) + k_p s(x)].$$

Here $\hat{g}(x)$ is an estimate of the bound generating function $g_\Gamma(x)$, $s(x)$ is the corresponding standard error, the critical value k_p is an estimate of the p-quantile of the maximum of the stochastic process

$$\mathbb{Z}_n(x) := \left(\frac{\hat{g}(x) - g_\Gamma(x)}{s(x)} \right),$$

and the sets $\hat{\mathcal{X}}^U$ and $\hat{\mathcal{X}}^L$ are both (random) subsets of the support of X that contain the points where the maximum and minimum is achieved with probablity tending to one, respectively. Following the recommendation in CLR, we set

$$\hat{\mathcal{X}}^U = \{x \in \mathcal{X} : \hat{g}_\Gamma(x) \geq \max_{x \in \mathcal{X}} \hat{g}_\Gamma(x) - 2\sqrt{\log(n)} \sup_{x \in \mathcal{X}} s(x)\}$$
$$\hat{\mathcal{X}}^L = \{x \in \mathcal{X} : \hat{g}_\Gamma(x) \leq \min_{x \in \mathcal{X}} \hat{g}_\Gamma(x) + 2\sqrt{\log(n)} \sup_{x \in \mathcal{X}} s(x)\}.$$

The specific choices of $\hat{g}$, s and k_p (and thus also those of $\hat{\mathcal{X}}^U$ and $\hat{\mathcal{X}}^L$) depend on the functional Γ of interest, and are explicitly described below for the case of the mean and the quantile functional. Finally, define the interval $\hat{\mathcal{A}}(p)$ as

$$\hat{\mathcal{A}}(p) = [\hat{\alpha}_p^L, \hat{\alpha}_p^U].$$

With this notation, the estimate of the identified set $\mathcal{A}(\Gamma, T)$ is then given by $\hat{\mathcal{A}}(1/2)$. In particular, using the choices described below, Theorem 1 in CLR implies that $\hat{\alpha}_{1/2}^U$ is a consistent and asymptotically median unbiased estimate of the upper bound $\alpha^U(\Gamma, T)$ of the identified set, in the sense that

$$\Pr(\alpha^U(\Gamma, T) \leq \hat{\alpha}_{1/2}^U) = 1/2 + o(1).$$

An analogous result applies for the lower bound. It is furthermore possible to construct two-sided confidence intervals for the true parameter value as follows: Let $\Delta_n^+ = \Delta_n \mathbb{I}\{\Delta_n > 0\}$, where $\Delta_n = \hat{\alpha}_{1/2}^U - \hat{\alpha}_{1/2}^L$, and $\hat{p}_n = \Phi(\tau_n \Delta_n^+)c$, where $\Phi(\cdot)$ is the standard normal CDF and $\tau_n = \log(n)/\max[\hat{\alpha}_{3/4}^U - \hat{\alpha}_{1/4}^U, \hat{\alpha}_{3/4}^L - \hat{\alpha}_{1/4}^L]$. Then $\hat{\mathcal{A}}(\hat{p}_n)$ provides an asymptotic $1 - c$ confidence interval for the parameter of interest, such that

$$\inf_{\alpha \in \mathcal{A}(\Gamma, T)} \Pr(\alpha \in \hat{\mathcal{A}}(\hat{p}_n)) \geq 1 - c + o(1).$$

These confidence intervals are thus valid uniformly with respect to the location of the true parameter value $\alpha(\Gamma, T)$ within the bounds. This follows from Theorem 3 in CLR.

B.2 Application to Mean and Quantile UPEs

In this section, we describe how to chose $\hat{g}$, s and k_p such that the conditions of Theorem 1 and 3 in CLR are satisfied, when $\Gamma(F_Y)$ is either the mean or some quantile of the outcome distribution. Other statistics of interest could be dealt with using similar arguments. Throughout this section, we assume that the following standard regularity conditions hold.

Assumption 1. *The sample observations $\{(Y_i, T_i, X_i)\}_{i=1}^n$ are a sequence of independent and identically distributed random vectors generated according to the model defined in Section 2.*

Assumption 2. *(i) The random vector $Z = (T, X')'$ has support $\{0,1\} \times \mathcal{X}$, where $\mathcal{X} = \{x_1, \ldots, x_r\} \subset \mathbb{R}^d$ is finite and consists of $r \geq 2$ elements. (ii) For every $(t, x) \in \{0,1\} \times \mathcal{X}$ the conditional variance $Var(Y|T = t, X = x) = \sigma^2(t, x)$ exists and is finite.*

Assumption 3. *The density function f_Y of Y is bounded away from zero around $Q_Y(\tau)$, is twice continuously differentiable, and the derivatives are uniformly bounded.*

Assumption 4. *The kernel function $K : \mathbb{R} \to \mathbb{R}$ satisfies (i) $\int K(y)dy = 1$, (ii) $\int yK(y)dy = 0$, (iii) $\int y^2 K(y)dy < \infty$, (iv) $\int K(y)^2 dy < \infty$, (v) K is Lipschitz continuous, (vi) $\int |K(y)|^{2+\mu} dy < \infty$, for some $\mu > 0$.*

B.2.1 Bounds on the Mean UPE

We start by consider the case where the functional of interest is the mean functional $\Gamma_M(F_Y) = \int y dF_Y(y)$. See Example 1 in Section 3 for details. Here our estimate of the bound generating function g_Γ is given by

$$\hat{g}(x) = \hat{\mathbb{E}}(Y|T = 1, X = x) - \hat{\mathbb{E}}(Y|T = 0, X = x),$$

where

$$\hat{\mathbb{E}}(Y|T = t, X = x) = \frac{1}{N(t, x)} \sum_{i=1}^{n} Y_i \mathbb{I}\{(T_i, X_i) = (t, x)\}$$

is the estimate of the conditional expectation of Y given T and X, and

$$N(t,x) = \sum_{i=1}^{n} \mathbb{I}\{(T_i, X_i) = (t,x)\}$$

is the number of observations within a cell defined by a realization of the covariate vector. The corresponding standard errors can then simply be calculated as

$$s(x) = \left(\frac{\hat{\sigma}^2(1,x)}{N(1,x)} + \frac{\hat{\sigma}^2(0,x)}{N(0,x)} \right)^{1/2}$$

where

$$\hat{\sigma}^2(t,x) = \frac{1}{N(t,x)} \sum_{i=1}^{n} Y_i^2 \mathbb{I}\{(T_i, X_i) = (t,x)\} - \hat{\mathbb{E}}(Y|T=t, X=x)^2$$

for $t = 0, 1$. Now since $\mathcal{X}$ is finite and $\hat{g}(x)$ is independent of $\hat{g}(\tilde{x})$ for $x \neq \tilde{x}$, it follows directly from Assumption 1–2 and the central limit theorem that

$$\mathbb{Z}_n(x) := \left(\frac{\hat{g}(x) - g_\Gamma(x)}{s(x)} \right) \stackrel{d}{=} \mathbb{Z}_\infty(x) + o_p(1) \quad \text{in } \ell^\infty(\mathcal{X}),$$

where $\mathbb{Z}_\infty(x)$ is a mean zero Gaussian process with $\mathrm{Var}(\mathbb{Z}_\infty(x)) = 1$ for all x and $\mathbb{Z}_\infty(x)$ being independent of $\mathbb{Z}_\infty(\tilde{x})$ for $x \neq \tilde{x}$. This implies that by Lemma 1 in CLR we can choose k_p as the p-quantile of $\mathbb{H}_\infty = \max_{x \in \hat{\mathcal{X}}} \mathbb{Z}_\infty(x)$. Due to the simple structure of $\mathbb{Z}_\infty$, this quantity is given by $k_p = \Phi^{-1}(p^{1/r})$, where r is the cardinality of $\hat{\mathcal{X}}^L$ or $\hat{\mathcal{X}}^R$, and $\Phi(\cdot)$ is the standard normal distribution function.

B.2.2 Bounds on the Quantile UPE

We now consider the case where the functional of interest is the quantile functional $\Gamma_{Q,\tau}(F_Y) = \inf\{y \in \mathbb{R} : F_Y(y) \geq \tau\} := Q_Y(\tau)$. See Example 2 in Section 3 for details. Our estimate of the bound generating function $g_\Gamma(x)$ is given by

$$\hat{g}(x) = -\frac{\hat{F}_{Y|T,X}(\hat{Q}_Y(\tau)|1,x) - \hat{F}_{Y|T,X}(\hat{Q}_Y(\tau)|0,x)}{\hat{f}_Y(\hat{Q}_Y(\tau))} \equiv -\frac{\hat{u}(\hat{Q}_Y(\tau), x)}{\hat{f}_Y(\hat{Q}_Y(\tau))}.$$

Here $\hat{Q}_Y(\tau)$ is the ordinary sample quantile of Y,

$$\hat{F}_{Y|T,X}(y,t,x) = \frac{1}{N(t,x)} \sum_{i=1}^{n} \mathbb{I}\{Y_i \leq y\} \mathbb{I}\{(T_i, X_i) = (t,x)\}.$$

is an estimate of the conditional CDF of Y given T and X, and

$$N(t,x) = \sum_{i=1}^{n} \mathbb{I}\{(T_i, X_i) = (t,x)\}$$

is again the number of observations within a cell defined by a realization of the covariate vector. Finally, $\hat{f}_Y$ is a kernel estimator of the density of Y, given by

$$\hat{f}_Y(y) = \frac{1}{n}\sum_{i=1}^{n} K_h(Y_i - y)$$

where $K_h(\cdot) = K(\cdot/h)/h$, K is a standard symmetric kernel function that integrates to one, and $h = h(n)$ is the bandwidth chosen such that as $h \to 0$, $nh \to \infty$ and $nh^5 \to 0$. For our empirical application, we use a Gaussian kernel and a slightly modified version of "Silverman's rule of thumb" to select the bandwidth, setting $h = 1.06\hat{\sigma}_Y n^{-1/4}$. The results are not sensitive to this choice.

For the construction of suitable standard errors, it is important to take into account that the different components of $\hat{g}$ converge to the corresponding true values at different rates: while $\hat{Q}_Y(\tau)$ and $\hat{F}_{Y|T,X}$ converge at the parametric rate $\sqrt{n}$, the density estimate $\hat{f}_Y$ is of the lower order $\sqrt{nh}$ and thus dominates the overall rate of convergence. From an asymptotic point of view, one could therefore act as if the former two quantities were in fact known, and compute standard errors that only account for the sampling variability of the density estimate. However, such an approach would be grossly misleading in our setting: Both $\hat{Q}_Y(\tau)$ and $\hat{f}_Y$ are computed from the entire sample of size $n = 266,956$, and are thus estimated very precisely. On the other hand, every value of the function $\hat{F}_{Y|T,X}(y,t,x)$ is computed only from the observations with $(T_i, X_i) = (t,x)$, which are less than 50 for many cells. We therefore use standard errors and a corresponding critical value that account for the substantial finite sample variability of $\hat{F}_{Y|T,X}$ through the inclusion of appropriate "higher order terms". In particular, we set

$$s(x) = \left(s_1(x)^2 + s_2(x)^2\right)^{1/2},$$

where

$$s_1(x) = \left(\frac{\hat{u}(\hat{Q}_Y(\tau), x)^2}{4nh\hat{f}_Y(\hat{Q}_Y(\tau))^3}\int K(v)^2 dv\right)^{1/2}$$

$$s_2(x) = \left(\frac{\hat{\sigma}_u^2(1,x)}{\hat{f}_Y(\hat{Q}_Y(\tau))^2 N(1,x)} + \frac{\hat{\sigma}_u^2(0,x)}{\hat{f}_Y(\hat{Q}_Y(\tau))^2 N(0,x)}\right)^{1/2}$$

and $\hat{\sigma}_u^2(t,x) = \hat{F}_{Y|T,X}(y,t,x)(1-\hat{F}_{Y|T,X}(y,t,x))$. Here $s_1(x)$ and $s_2(x)$ are the contributions of estimating f_Y and $F_{Y|T,X}$, respectively, to the overall standard error $s(x)$. Regarding the critical value, we set

$$k_p = \Phi^{-1}(p) \max_{x\in\hat{\mathcal{X}}} \frac{s_1(x)}{s(x)} + \Phi^{-1}(p^{1/r}) \min_{x\in\hat{\mathcal{X}}} \frac{s_2(x)}{s(x)},$$

where r is the cardinality of $\hat{\mathcal{X}}^L$ or $\hat{\mathcal{X}}^R$, and $\Phi(\cdot)$ is the standard normal CDF.

For the data used in our empirical application, we have that $s_1(x) \approx 0$ for all $x \in \mathcal{X}$ and all τ being considered, so that $s(x) \approx s_2(x)$ and $k_p \approx \Phi^{-1}(p^{1/r})$. Our choices thus essentially correspond to the case where $Q_Y(\tau)$ and f_Y are known, which is completely analogous to the case of the mean described in the previous subsection. On the other hand, our choices are asymptotically valid and satisfy the conditions of Theorem 1 in CLR. To see this, we can use Assumption 1 and 3–4 together with the usual Taylor expansion arguments and write

$$\begin{aligned}\hat{g}(x) - g_\Gamma(x) &= \frac{u(Q_Y(\tau),x)}{2f_Y(Q_Y(\tau))^2}(\hat{f}_Y(Q_Y(\tau)) - f_Y(Q_Y(\tau))) + o_p((nh)^{-1/2}) \\ &\stackrel{d}{=} \mathcal{N}\left(0, \frac{u(Q_Y(\tau),x)^2}{4f_Y(Q_Y(\tau))^3}\int K(v)^2 dv\right) + o_p(1)\end{aligned}$$

for each $x \in \mathcal{X}$. Furthermore, we have that $s(x) = s_1(x) + o_p(s_1(x))$. It then follows from the Central Limit Theorem that

$$\mathbb{Z}_n(x) := \left(\frac{\hat{g}(x) - g_\Gamma(x)}{s(x)}\right) = \left(\frac{\hat{g}(x) - g_\Gamma(x)}{s_1(x)}\right) + o_p(1) \stackrel{d}{=} \mathbb{Z}_\infty + o_p(1) \quad \text{in } \ell^\infty(\mathcal{X}),$$

where $\mathbb{Z}_\infty = \mathcal{N}(0,1)$ is simply a standard normal random variable that does not depend on x. Hence, by Lemma 1 in CLR any critical value k_p equal to $\Phi^{-1}(p) + o_p(1)$ satisfies the conditions of their Theorem 1. In particular, our choice of $k_p = \Phi^{-1}(p)(1 + O_p(h^{1/2})) + O_p(h^{1/2}) = \Phi^{-1}(p) + O_p(h^{1/2})$ is valid.

References

Chernozhukov, V., I. Fernandez-Val, and B. Melly (2009): "Inference on Counterfactual Distributions," *Working Paper.*

Chernozhukov, V., S. Lee, and A. Rosen (2009): "Intersection Bounds: Estimation and Inference," *Working Paper.*

DiNardo, J., N. Fortin, and T. Lemieux (1996): "Labor market institutions and the distribution of wages, 1973-1992: A semiparametric approach," *Econometrica*, 64(5), 1001–1044.

Donald, S., D. Green, and H. Paarsch (2000): "Differences in wage distributions between Canada and the United States: An application of a flexible estimator of distribution functions in the presence of covariates," *Review of Economic Studies*, 67(4), 609–633.

Firpo, S., N. Fortin, and T. Lemieux (2007): "Unconditional quantile regressions," *NBER Working Paper*.

——— (2009): "Unconditional quantile regressions," *Econometrica*, 77(3), 953–973.

Joe, H. (1997): *Multivariate models and dependence concepts*. Chapman & Hall/CRC.

Machado, J., and J. Mata (2005): "Counterfactual decomposition of changes in wage distributions using quantile regression," *Journal of Applied Econometrics*, 20(4).

Nelsen, R. (2006): *An introduction to copulas*. Springer Verlag.

Rothe, C. (2009): "Nonparametric Estimation of Distributional Policy Effects," *Working Paper*.

Sklar, A. (1959): "Fonctions de répartition à n dimensions et leurs marges," *Publ. Inst. Statist. Univ. Paris*, 8, 229–231.

Trivedi, P., and D. Zimmer (2007): *Copula modeling: an introduction for practitioners*. Now Publishers Inc.

Van der Vaart, A. (2000): *Asymptotic Statistics*. Cambridge University Press.

Wooldridge, J. (2002): *Econometric analysis of cross section and panel data*. MIT Press.