Partial Distributional Policy Effects

Christoph Rothe*

Toulouse School of Economics

Preliminary version. Comments are welcome.

Abstract

In this paper, we propose a method to evaluate the effect of a counterfactual change in the marginal distribution of a single covariate on the unconditional distribution of an outcome variable of interest. Both fixed and infinitesimal changes are considered. We show that such effects are point identified under general conditions if the covariate affected by the counterfactual change is continuously distributed, but are typically only partially identified if its distribution is discrete. For the latter case, we derive informative bounds making use of the available information. We also discuss estimation and inference.

JEL Classification: C14, C31

Keywords: *Counterfactual Distribution, Partial Identification*

*This Version: October 14, 2010. Address for correspondence: Toulouse School of Economics (GREMAQ), 21 Allée de Brienne, 31000 Toulouse, France. E-Mail: rothe@cict.fr. Website: http://www.christophrothe.net. This is a heavily revised version of material which previously circulated under the titles "Decomposing Counterfactual Distributions" and "Marginal Distributional Policy Effects of Binary Covariates". I would like to thank Joshua Angrist, Andrew Chesher, Sergio Firpo, Thomas Lemieux, Thierry Magnac and conference participants at the ESEM 2009 Barcelona and EALE/SOLE 2010 Meeting London for helpful comments. The usual disclaimer applies.

1. Introduction

In this paper, we propose a method to evaluate the effect of a counterfactual change in the unconditional distribution of a *single* component of a vector of explanatory variables X on the unconditional distribution of an outcome variable Y, holding all other features of the system that relates Y and X constant. Both fixed and marginal (infinitesimal) counterfactual changes can be considered. Such unconditional *ceteris paribus* effects are of interest in many areas of applied policy analysis. For example, one might be interested what the distribution of wages in 2010 would be if workers' age structure was still as in 1980, but all other characteristics of the labor force were the same as today; or how the distribution of wages would react to a marginal increase in the proportion of unionized workers, holding again all other characteristics of the labor force constant. These questions are straightforward to address for the mean of Y in a linear model. If $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$, and we write $\mathbb{E}(X_1) = \mu_1$, then mechanically a change in the distribution of X_1 to another one with mean μ^* increases the expectation of Y by $(\mu^* - \mu_1)\beta_1$. This is the main idea behind the popular Oaxaca-Blinder procedure (Oaxaca, 1973; Blinder, 1973) to decompose intra-group differences in means. On the other hand, the effect of a marginal increase in μ_1 is easily seen to be equal to $\partial\mathbb{E}(Y)/\partial\mu_1 = \beta_1$.

The main contribution of this paper is to introduce a class of parameters called *Partial Distributional Policy Effects (PPEs)* that generalize these ideas in three important directions: First, they allow for general, both fixed and marginal, changes in one of the covariates' distribution, and not only location shifts. Second, they allow for arbitrarily complex nonlinear relationships between the outcome variable and the covariates, instead of relying on the linear model. And third, they measure the impact on general distributional features of the distribution Y, such as its variance, quantiles or Gini coefficient, and not only on the mean. The paper thus extends earlier work on distributional policy effects in Rothe (2010), which considers the effect of fixed changes in the *entire* covariate distribution. Building on the literature on nonseparable models (e.g. Matzkin, 2003), we formalize the ambiguous notion of a *ceteris paribus* change in one of the components of the covariate distribution by imposing a rank invariance condition. That is, we construct the counterfactual experiment in such a way that the joint distribution of covariates' ranks remains unaffected. This is equivalent to holding the copula function of the covariate distribution constant, and thus preserves the dependence structure. We show that under a conditional exogeneity condition the distribution of the counterfactual outcome variable can be obtained by integrating the conditional CDF of Y given X with respect to the new counterfactual covariate distribution, and one can thus directly calculate distributional features of interest. We also discuss both parametric and nonparametric sample analogue estimators based on this result.

A particular complication arises for discrete covariables. In this case, the rank of an individual in the respective unconditional distribution is not uniquely determined by the data. Due to this particular form of interval-censoring, the corresponding policy parameters are typically only partially identified. That is, the data generating process reveals some nontrivial information about these effects, but does not allow for an exact quantification. This finding should not be seen as a weakness of our approach, but points to the difficulties to define such effects in general nonlinear models. Following the literature on partially identified parameters (e.g. Manski, 2003, 2007), we derive bounds on the PPEs in this case.

This paper contributes to an extensive literature on the analysis of counterfactual distributions. The impact of fixed changes in the *entire* covariate distribution is studied for example by Stock (1989), DiNardo, Fortin, and Lemieux (1996), Gosling, Machin, and Meghir (2000), Donald, Green, and Paarsch (2000), Barsky, Bound, Charles, and Lupton (2002), Machado and Mata (2005), Melly (2005), Chernozhukov, Fernandez-Val, and Melly (2009a) and Rothe (2010). For the special case of a dummy variable, DiNardo et al. (1996) also propose a reweighting procedure to quantify the the effect of a change in the *conditional* distribution of one of the covariates given the remaining covariates. Machado and Mata (2005) consider a related reweighting scheme, which alters both the dependence structure and the unconditional distribution of remaining covariates. None of these papers provides a general method to quantify the effect of changes in the *unconditional* distribution of a *single* explanatory variable, holding everything else constant. In a framework similar to ours, Firpo, Fortin, and Lemieux (2009) consider the impact of marginal location shifts in continuously distributed covariates, and of marginal changes in the *conditional* distribution of a binary covariate given the remaining covariates via their so-called RIF-regression approach. While the former parameter is a special case of our PPEs, the latter is substantially different from the unconditional effects we consider in this paper. A similar comment applies to the Marginal Policy-Relevant Treatment Effect studied by Carneiro, Heckman, and Vytlacil (2010), which corresponds to the effect of a marginal change in the *conditional* probability of receiving a (binary) treatment given a vector of instruments. The aim of our paper is also similar to that of Firpo, Fortin, and Lemieux (2007; 2010), who, using their RIF-regression techniques, propose a procedure to determine the effect of changes in the distribution of a single covariate on general features of the outcome distribution. Their approach relies on a particular linearity restriction on RIF-regression function, which implies that the respective distributional feature of the outcome variable depends on the distribution of the covariates only through its vector of means. Such a condition does not hold for common distributional features other than the mean, e.g. quantiles or variances, even for simple data generating processes like linear models. In contrast, our paper does not impose shape restrictions on the conditional distribution of the outcome given the explanatory variables.

The remainder of the paper is organized as follows. In the next section, we introduce our model and the parameters of interest. Section 3 contains the identification analysis, and Section 4 discusses estimation and inference. Section 5 concludes. All proofs are collected in Appendix A. Further details are discussed in Appendix B–E.

2. Model and Parameters of Interest

We observe an outcome variable Y and a d-dimensional vector of covariates X, which are related through a general nonseparable structural model

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

where $\eta \in \mathbb{R}^{d_\eta}$ is an unobserved error term. Since we do not impose any restrictions on neither the dimension of the unobservables nor the way they enter the structural function m, the model in (2.1) allows for flexible forms of unobserved heterogeneity. In the following, we index distribution and quantile functions by the random variables they refer to, so that F_Y denotes the CDF of Y, etc.

Our aim in this paper is to study the effect of a counterfactual (fixed or marginal) change in the unconditional distribution of one of the covariates on some feature of the distribution of Y, holding everything else, in particular the dependence structures and the distribution of the remaining covariates constant. To formalize the ambiguous notion of a *ceteris paribus* change in one of the components of a multivariate distribution, we partition the covariate vector as $X = (W, Z)$, where W is the one-dimensional random variable whose unconditional distribution is going to be changed in the counterfactual experiment, and Z is the d'-dimensional vector of remaining covariates. We then write the observed covariates X in terms of their marginal quantile functions and a vector $U = (U_1, \dots, U_d)$ of standard uniformly distributed latent variables, i.e.

$$X = (Q_W(U_1), Q_{Z_1}(U_2), \dots, Q_{Z_{d'}}(U_d)) \tag{2.2}$$

for some $U_i \sim U[0,1]$ and $i = 1, \dots, d$. We refer to U in the following as the vector of rank variables, and denote its joint CDF, which is also the copula function of F_X, by C. If W is continuously distributed, the latent rank variable U_1 constitutes a one-to-one transformation of W, since the quantile function Q_W is strictly increasing and thus injective in this case. If W is binary, e.g. an indicator of union membership, the relationship $W = Q_W(U_1) = \mathbb{I}\{U_1 > \Pr(W = 0)\}$ can be thought of as a threshold crossing model. However, it is important to stress that (2.2) is not a "model", but simply a representation that can be assumed without loss of generality.

It is evident from (2.2) that the quantile functions only determine the shape of the marginal distributions of X, whereas the vector of rank variables U determines its dependence structure. It is therefore natural to define the outcome Y_H of the counterfactual experiment in which the unconditional distribution of W has been changed to some CDF H, but everything else has been held constant, as

$$Y_H = m(X_H, \eta),$$

where $X_H = (H^{-1}(U_1), Q_{Z_1}(U_2), \dots, Q_{Z_{d'}}(U_d)) = (H^{-1}(U_1), Z)$ is the corresponding counterfactual covariate vector, and $H^{-1}(\tau) = \inf\{w : H(w) \geq \tau\}$ is the quantile function corresponding to H. Depending on the application, H could either be a fixed CDF, such as the distribution of W in a different population, or part of a sequence of CDFs

$\{H_t, t \in \mathbb{R}\}$ that tends to F_W from a particular direction as $t \to 0$. Note that our definition is equivalent to imposing a rank invariance condition, since the unconditional distribution of W is changed in such a way that the joint distribution of ranks of X remains unaffected.

Our aim is to learn about various features $\nu(F_Y^H)$ of the distribution F_Y^H of Y_H, and to compare them to the corresponding features $\nu(F_Y)$ of the distribution of Y. We refer to any difference between these quantities as a *Partial Distributional Policy Effect (PPE).* Here $\nu : \mathcal{F} \to \mathbb{R}$ is a functional from the space of all one-dimensional distribution functions to the real line. One example for such a feature would be the mean of Y_H, which can be written as $\mathbb{E}(Y_H) = \mu(F_Y^H)$ for $\mu : F \mapsto \int y dF(y)$. Other examples are higher-order centered or uncentered moments, quantiles and related statistics like interquantile ranges or quantile ratios, and inequality measures such as the Gini coefficient.[1] Our parameters of interest are formally defined as follows.

Definition 1. *(a) Let H be a fixed CDF. Then the* Fixed Partial Distributional Policy Effect (FPPE) *is given by*

$$\alpha_W(\nu) = \nu(F_Y^H) - \nu(F_Y).$$

(b) Let $H = H_t$ be an element of a continuum of CDFs indexed by $t \in \mathbb{R}$ such that $H_t \to F_W$ as $t \to 0$, and denote the CDFs of the corresponding counterfactual outcome distributions by F_Y^t. Then the Marginal Partial Distributional Policy Effect (MPPE) *is given by*

$$\beta_W(\nu) = \lim_{t \to 0} \frac{\nu(F_Y^t) - \nu(F_Y)}{t} = \partial_t \nu(F_Y^0),$$

provided that the limit exists.

Following Firpo et al. (2009), we will focus on MPPEs corresponding to either marginal location shifts $H_t(w) = F_W(w-t)$ or marginal perturbations $H_t(w) = F_W(w)+t(G_W(w)-F_W(w))$ in some fixed direction G_W.

[1] See Chernozhukov et al. (2009a) or Rothe (2010) for further examples and an extensive discussion.

3. Identification

Following the literature on counterfactual distributions, we establish our identification results assuming a form of conditional exogeneity (e.g. Firpo et al., 2009; Chernozhukov et al., 2009a; Rothe, 2010). All issues concerning identification that we address in this paper are not specific to this setting, but would appear in the same fashion in a nonseparable model with endogeneity (e.g. Chesher, 2003; Imbens and Newey, 2009). We first obtain two useful representations of the counterfactual outcome distribution F_Y^H.

Lemma 1. *Assume that (a) the unobserved heterogeneity η is independent of U_1 conditional on Z, i.e. $\eta \perp U_1 | Z$, and (b) the support of H is a subset of the support of W conditional on Z, i.e. $supp(H) \subset supp(W|Z = z)$ for all $z \in supp(Z)$. Then the counterfactual outcome distribution F_Y^H can be either written as (i) $F_Y^H(y) = \mathbb{E}(F_{Y|X}(y|H^{-1}(U_1), Z))$, or as (ii) $F_Y^H(y) = \int F_{Y|X}(y|w, z)dC(H(w), F_{Z_1}(z_1), \ldots, F_{Z_{d'}}(z_{d'}))$.*

Condition (a) of the Lemma is sufficient but not necessary for conditional exogeneity of $W = Q_W(U_1)$ if W is discrete, and equivalent to conditional exogeneity if W is continuously distributed. It is substantially weaker than assuming full exogeneity of X. In the context of identification in nonseparable models, a similar assumption is employed by Hoderlein and Mammen (2007). Condition (b) ensures that the support of X_H is a subset of the support of X, and thus that the function $F_{Y|X}$ is identified over the area of integration. Since we treat the structural function m in a nonparametric fashion, extrapolation outside the range of observed covariates is not possible in our setting.

Lemma 1 shows that identification of our PPEs hinges upon knowledge of the rank variable U_1 or, equivalently, the copula function C. Such knowledge is not available when W is discrete, and thus there are generally several CDFs that can be written as $F(y) = \mathbb{E}(F_{Y|X}(y|H^{-1}(\tilde{U}_1), Z))$ for some $\tilde{U}_1 \sim U[0, 1]$ such that $(Q_W(\tilde{U}_1), Z) \stackrel{d}{=} (W, Z)$, and can thus not be ruled out as possible values of F_Y^H from the data. We denote the set of all such *feasible counterfactual outcome distributions* by $\mathcal{F}_Y^H$. Note that this set could be defined equivalently via the copula representation in Lemma 1(ii).

3.1. Fixed Partial Policy Effects. When W is continuously distributed, the quantile function Q_W is strictly increasing, and establishes a one-to-one relationship between

$W = Q_W(U_1)$ and the latent rank variable U_1. Thus, by Lemma 1 the counterfactual outcome distribution F_Y^H is point identified, as are of course all distributional features of the form $\nu(F_Y^H)$, and thus the FPPE. The next theorem formalizes this result.

Theorem 1. *Assume the conditions of Lemma 1 hold. Then we have that* $F_Y^H(y) = \mathbb{E}(F_{Y|X}(y|H^{-1}(F_W(W)), Z)))$*, and the FPPE* $\alpha_W(\nu)$ *is identified for any functional* ν*.*

When W is not continuously distributed, the quantile function Q_W is piecewise constant, and it can thus only be deduced from observing W that $F_W(W-) < U_1 \leq F_W(W)$, where the notation $f(x-)$ denotes the left limit of the function f at the point x. Interval-censoring of covariates is well-known to lead to identification problems in various contexts (Manski and Tamer, 2002), and prevents point identification of the FPPE in our context. In order to derive the identified set, we make use of the following lemma.

Lemma 2. *For every* $F \in \mathcal{F}_Y^H$ *there exists a random variable* V *called a* rank allocator *satisfying the relationship* $V|W = w \sim U[0,1]$ *for all* $w \in supp(W)$*, such that* $F(y) = \mathbb{E}(F_{Y|X}(y, H^{-1}(\tilde{F}_W(W,V)), Z))$*, with* $\tilde{F}_W(w,v) = v(F_W(w) - F_W(w-)) + F_W(w-)$*.*

The role of the rank allocator, which is allowed to depend on Z, is to assign a unique rank to each individual in case that F_W is not continuous everywhere. The idea behind this construction is that since we only know that $F_W(W-) < U_1 \leq F_W(W)$, all uniform allocations of ranks within these bounds are observationally equivalent, and thus lead to a feasible value of the counterfactual outcome distribution.

Using Lemma 2, we now derive sharp bounds on $\nu(F_Y^H)$ for linear functionals ν by explicitly constructing appropriate rank allocators. For simplicity, we focus on the important special case that F_W and H are supported on $\{0,1\}$. To illustrate the idea, suppose for the moment that $p_1 := \Pr(W = 1) > \Pr(H^{-1}(U_1) = 1) =: p_2$. Roughly speaking, this means that we have to "move" a fraction of $(p_1 - p_2)/p_1$ of the p_1 individuals with $W = 1$ to the group with $W = 0$ in the counterfactual experiment. When ν is linear, upper and lower bounds on $\nu(F_Y^H)$ can then be obtained by ranking all individuals with $W = 1$ according the individual effect $\nu(F_{Y|X}(\cdot|1, Z)) - \nu(F_{Y|X}(\cdot|0, Z))$ of such a "move", and selecting those at the top and the bottom of the ranking, respectively. More specifically, Lemma 2 and linearity of ν implies that for any $F \in \mathcal{F}_Y^H$ we have that

$$\begin{aligned}\nu(F) = \mathbb{E}[\nu(F_{Y|X}(\cdot|\mathbb{I}\{V > (p_1 - p_2)/p_1\}, Z))|W = 1]p_1 \\ + \mathbb{E}[\nu(F_{Y|X}(\cdot|0, Z))|W = 0](1 - p_1),\end{aligned} \tag{3.1}$$

for some rank allocator V.[2] The second term on the right-hand side of (3.1) does not depend on V and can thus be neglected. Depending on the realization of V, the term $\nu(F_{Y|X}(\cdot|\mathbb{I}\{V > (p_1-p_2)/p_1\}, Z))$ is equal to either $\nu(F_{Y|X}(\cdot|1, Z))$ or $\nu(F_{Y|X}(\cdot|0, Z))$. In order to maximize the right-hand side of (3.1), the rank allocator must thus be defined in such a way conditionally on $W = 1$ the event $V < (p_1-p_2)/p_1$ corresponds to a realization of $\tilde{V} = \nu(F_{Y|X}(\cdot|1, Z)) - \nu(F_{Y|X}(\cdot|0, Z))$ below its conditional $(p_1 - p_2)/p_1$-quantile. This can be achieved by defining V as an appropriately normlized version of $\tilde{V}$. A lower bound on the expression in (3.1) can be constructed by replacing $\tilde{V}$ by its negative version and proceeding analogously. In general, we thus first define the random variable

$$\tilde{V}_\nu = \nu(F_{Y|X}(\cdot|H^{-1}(F_W(W)), Z)) - \nu(F_{Y|X}(\cdot|H^{-1}(F_W(W-)), Z)).$$

Next, let V_ν^U be a one-to-one transformation of $\tilde{V}_\nu$ normalized to be standard uniformly distributed conditional on W via the (generalized) distributional transform,[3] and define $V_\nu^L = 1-V_\nu^U$. Then the two CDFs built from the rank allocators V_ν^L and V_ν^U, respectively, are those which yield the lowest and highest feasible value of $\nu(F_Y^H)$.

Theorem 2. *Suppose that the conditions of Theorem 1 hold, and let ν be a linear functional. Then $\alpha_W^L(\nu) \leq \alpha_W(\nu) \leq \alpha_W^U(\nu)$, where $\alpha_W^r(\nu) = \nu(F^r) - \nu(F_Y)$ and $\nu(F^r) = \mathbb{E}(\nu(F_{Y|X}(\cdot|S_\nu^r(W, Z), Z)))$ with $S_\nu^r(W, Z) = H^{-1}(\tilde{F}_W(W, V_\nu^r))$ for $r \in \{U, L\}$. In the absence of further information, these bounds are sharp.*

When the structural function satisfies a separability condition of the form $m(w, z, e) = m^A(w, e)+m^B(z, e)$, the upper and lower bound coincide since ν is linear, and the FPPE

[2] The last equality follows from the fact that for $p_2 < p_1$ we have that $Q_W(u) = \mathbb{I}\{u > 1 - p_1\}$, $H^{-1}(u) = \mathbb{I}\{u > 1 - p_2\}$, $\tilde{F}_W(0, V) = V(1 - p_1) < 1 - p_2$ and $\tilde{F}_W(1, V) = (1 - p_1) + Vp_1$.

[3] A random variable Q is said to be a normalized version of a random variable R conditional on a random vector S on via the generalized distributional transform if $Q = G(R, S, T)$, where $G(r, s, t) = \Pr(R < r|S = s) + t \Pr(R = r|S = s)$ and $T \sim U[0, 1]$ is some random variable independent of (R, S). See Rüschendorf (2009) for details.

is thus point identified irrespective of whether W is continuously distributed or not. Hence there are for example no identification issues related to discrete covariables in the classical Oaxaca-Blinder procedure, which is based on the linear model $m(w, z, e) = \alpha + \beta w + \gamma' z + e$. Furthermore, one can show that the Theorem also holds in the non-binary case, assuming that the function H^{-1} takes at most two values on the interval $(F_W(w-), F_W(w)]$ for every $w \in \mathbb{R}$.

With the exception of the mean, most distributional features commonly used in empirical applications cannot be written as linear functionals of the underlying CDF. As a first step to extend the result in Theorem 2, note that for every fixed $y \in \mathbb{R}$ the mapping $F \mapsto F(y)$ is linear. The value $F_Y^H(y)$ can thus be bounded pointwise using the approach described above, by constructing appropriate rank allocators V_y^L and V_y^U depending on $y \in \mathbb{R}$. Let $G^U(y)$ and $G^L(y)$ be the corresponding lower and upper bounds, respectively. Then we have that

$$G^U(y) \leq F_Y^H(y) \leq G^L(y) \text{ for all } y \in \mathbb{R}. \tag{3.2}$$

Since the rank allocation schemes V_y^L and V_y^U depend on the point of evaluation y, the functions G^L and G^U are not necessarily feasible counterfactual outcome distributions themselves, and thus constitute only pointwise but not uniformly sharp bounds. However, one can show that both are proper distribution functions that constitute best possible bounds on F_Y^H with respect to the partial ordering induced by first-order stochastic dominance. Using results in Stoye (2010), who derives identification regions for a large class of distributional features when the underlying CDF is restricted by first-order stochastic dominance bounds, we then obtain bounds on the FPPE if ν is either a D_1-parameter (e.g. mean, median, fixed quantile), a D_2-parameter (e.g. variance, Gini coefficient, Theil's index, Lorenz share), or a quantile contrast (e.g. interquantile range). To state the bounds, we also require the notion of compressed and dispersed distributions, which are those CDFs satisfying (3.2) that allocate as much probability mass as possible to their center and the tails, respectively. Exact definitions of the just-mentioned concepts are given in Appendix B.

Theorem 3. *Assume the conditions of Theorem 1 hold, and that ν is either a D_1-parameter, D_2-parameter, or quantile contrast. Then $\alpha_W^L(\nu) \leq \alpha_W(\nu) \leq \alpha_W^U(\nu)$, where* *the upper and lower bounds are given as follows:*

(i) For ν a D_1-parameter, we have $\alpha^r_W(\nu) = \nu(G^r) - \nu(F_Y)$ for $r \in \{U, L\}$.

(ii) For ν a D_2-parameter, assume that $\mu(F_Y^H) = \bar{\mu}$ for some $\bar{\mu} \in (\mu_H^L, \mu_H^U)$, and let $G^U_{\bar{\mu}}$ and $G^L_{\bar{\mu}}$ be the unique compressed and dispersed distributions (relative to G^U and G^L) with expectation $\mu(G^L_{\bar{\mu}}) = \mu(G^U_{\bar{\mu}}) = \bar{\mu}$. Then we have $\alpha^r_W(\nu) = \nu(G^r_{\bar{\mu}}) - \nu(F_Y)$ for $r \in \{U, L\}$.

(iii) For ν an (α, β)-quantile contrast, choose any $\gamma \in (\alpha, \beta)$ and $\bar{m} \in (G^L(\gamma), G^U(\gamma))$, let $G^L_{\bar{m}}$ be the compressed distribution with threshold value $a = \bar{m}$, and G^U_{γ} be the dispersed distribution with threshold value $a = \gamma$. Then we have $\alpha^L_W(\nu) = \nu(G^L_{\bar{m}}) - \nu(F_Y)$ and $\alpha^U_W(\nu) = \nu(G^U_{\gamma}) - \nu(F_Y)$.

Since the functions G^L or G^U may not be feasible values of F_Y^H themselves, the bounds in Theorem 3 may not be sharp.[4] It should in principle be possible to tighten these bounds by tailoring the construction of the rank allocator to the respective functional of interest. However, except for the important special case of linear parameters discussed in detail above, there seems to be no straightforward analytical solution to this problem. Note that the result in Theorem 3(ii) does not require the mean of F_Y^H to be point identified (which would generally not be the case in our setting). Instead, together with Theorem 2 it establishes a joint identification region for the mean and any D_2-parameter, whose shape is typically not rectangular.

3.2. Marginal Partial Policy Effects. Using the copula representation of the counterfactual outcome distribution in Lemma 1(ii), it is easy to see that for continuously distributed W the MPPE is identified if the functional ν and the copula function C satisfy appropriate smoothness conditions. The result is analogous to findings in Firpo et al. (2009), and stated for completeness.

[4] Stoye (2010) notes that even if G^L and G^U are feasible distributions of Y_H, one could possibly improve upon the above bounds if there is additional information available about the structural function m, that e.g. implies that Y_H is discrete. Such additional information could be easily included in our analysis at the price of a substantially more involved notation.

Theorem 4. *Suppose that (a) the conditions of Theorem 1 hold with H replaced by H_t for all $t \in \mathbb{R}$ sufficiently close to zero, (b) ν is Hadamard differentiable*[5] *at F_Y with derivative ν', (c) the partial derivative $\partial C/\partial u_1 = C_1$ of the copula function C exists, and (d) the unconditional distribution of W is continuous. Then the MPPE $\beta_W(\nu)$ is identified:*

(i) For H_t a marginal perturbation, i.e. $H_t(w) = (1-t)F_W(w) + tG_W(w)$, we have

$$\beta_W(\nu) = \int \nu'(F_{Y|X}(y|w,z))d(C_1(F_W(w), F_Z(z))(G_W(w) - F_W(w))).$$

(ii) For H_t a marginal location shift, i.e. $H_t(w) = F_W(w-t)$, we have

$$\beta_W(\nu) = \nu'\left(\mathbb{E}(\partial_w F_{Y|X}(\cdot|W,Z))\right).$$

When W is not continuously distributed, a marginal location shift does not satisfy the support condition in Lemma 1, and hence we only consider marginal perturbations in this case. While the formula in Theorem 4(i) remains valid, the MPPE is typically no longer point identified for discrete W, as one is unable to learn the partial derivative of the copula function in this context. To see this, we focus on the case that W is supported on $\{0,1\}$, and consider a perturbation H_t which implies an increase in the probability of observing $W = 1$ by t, i.e. $H_t(w) = \mathbb{I}\{0 \leq t < 1\}(F_W(0) - t) + \mathbb{I}\{t \geq 1\}$. Then it follows from direct calculations that

$$\beta_W(\nu) = \int (\nu'(F_{Y|X}(\cdot|1,z)) - \nu'(F_{Y|X}(\cdot|0,z)))dC_1(F_W(0), F_Z(z)).$$

By Sklar's Theorem (Sklar, 1959; Nelsen, 2006, Theorem 2.3.3), the copula C is only identified on the range of the marginal CDFs of $X = (W,Z)$. When W is binary, the function $C(a,\cdot)$ is thus identified for $a \in \{0, F_W(0), 1\}$ only. This in turn implies that the function $C_1(F_W(0),\cdot)$ is not point identified, since identification of a derivative at a fixed point requires knowledge of the function at least in some small neighborhood. In order to still obtain bounds on the MPPE, we show in the Appendix that the set of all possible values of the function $C_1(F_W(0),\cdot)$ that are compatible with the distribution of observables is the set of all multivariate distribution functions with support $R_{\mathcal{Z}} =$

[5] A formal definition of Hadamard differentiability is given in Appendix B.

$\{(F_{Z_1}(z_1), \ldots, F_{Z_{d'}}(z_{d'})) : z \in \mathcal{Z}\}$, where $\mathcal{Z}$ denotes the support of Z. The identified set of the MPPE is thus the set of all density-weighted averages of the function

$$g_\nu(z) = \nu'(F_{Y|X}(\cdot|1, z)) - \nu'(F_{Y|X}(\cdot|0, z)),$$

and sharp upper and lower bounds are thus given by the extrema of this function over the support of Z.

Theorem 5. *Suppose that the conditions (i)–(iii) of Theorem 4 hold. Then we have that* $\beta_W^L(\nu) \leq \beta_W(\nu) \leq \beta_W^U(\nu)$, *where* $\beta_W^U(\nu) = \sup_{z\in\mathcal{Z}} g_\nu(z)$ *and* $\beta_W^L(\nu) = \inf_{z\in\mathcal{Z}} g_\nu(z)$.

It is an immediate consequence of Theorem 5 that $\beta_W(\nu)$ is identified if and only if $g_\nu(z)$ is constant for all $z \in \mathcal{Z}$. This would e.g. be the case if $\nu = \mu$ is the mean functional, which implies that $g_\nu(z) = \mathbb{E}(Y|W = 1, Z = z) - \mathbb{E}(Y|W = 0, Z = z)$, and the structural function m satisfies the separability condition $m(w, z, e) = m^A(w, e) + m^B(z, e)$. In contrast to the FPPE however, the separability condition alone is not sufficient to obtain point identification of the MPPE for distributional features other than the mean, such as e.g. quantiles: if $\nu(F) = F^{-1}(\tau)$ we have that $g_\nu(z) = -(F_{Y|X}(Q_Y(\tau)|1, z) - F_{Y|T,X}(Q_Y(\tau)|0, z))/f_Y(Q_Y(\tau))$, which generally varies with z.

4. Estimation and Inference

In this section, we discuss both parametric and nonparametric estimation of partial policy effects. Under point identification, both FPPEs and MPPEs can be estimated by simple "plug-in" procedures, replacing unknown quantities in the respective expressions in Theorem 1 and 4 with suitable sample counterparts. Under partial identification, estimates of the identified set can be obtained through "plug-in" estimates of the respective boundaries for FPPEs, and via the approach in Chernozhukov, Lee, and Rosen (2009b) for MPPEs.

4.1. Fixed Partial Policy Effects. We assume that the data consist of an i.i.d. sample of size n, i.e. we observe $(Y_i, W_i, Z_i)_{i=1}^n$. For many applications, such as Oaxaca-Blinder-type decompositions, the counterfactual covariate distribution H is not going to be known exactly, but estimated from a sample $(W_i^*)_{i=1}^{n^*}$ of size $n^* = n/\lambda$ for some $\lambda > 0$.

When W is continuously distributed, the identification result in Theorem 1 suggests to estimate the FPPE by

$$\hat{\alpha}_W(\nu) = \nu(\hat{F}_Y^H) - \nu(\hat{F}_Y).$$

Here $\hat{F}_Y^H(y) = (1/n)\sum_{i=1}^n \hat{F}_{Y|X}(y, \hat{H}^{-1}(\hat{F}_W 1(W_i)), Z_i)$, where $\hat{F}_Y$, $\hat{F}_W$ and $\hat{H}^{-1}$ denote the empirical CDF and quantile function of Y, W and W^*, respectively, and $\hat{F}_{Y|X}$ is an estimate of the conditional CDF of Y given X. The latter can be estimated by either of the parametric methods discussed in Chernozhukov et al. (2009a), e.g. by first estimating a linear quantile regression model $Q_{Y|X}(\tau, x) = x'\beta(\tau)$, and then inverting the corresponding conditional quantile quantile function, or by a fully nonparametric CDF estimator, e.g. a kernel estimator as in Rothe (2010).

Under partial identification, we obtain estimates $\hat{\mathcal{A}}_W(\nu)$ of the identified set of the FPPE by estimating the respective upper and lower boundaries, i.e. we have that $\hat{\mathcal{A}}_W(\nu) = [\hat{\alpha}_W^L(\nu), \hat{\alpha}_W^U(\nu)]$, where

$$\hat{\alpha}_W^r(\nu) = \begin{cases} \nu(\hat{F}^r) - \nu(\hat{F}_Y) & \text{if } \nu \text{ is a linear functional,} \\ \nu(\hat{G}^r) - \nu(\hat{F}_Y) & \text{if } \nu \text{ is a } D_1 \text{ parameter,} \\ \nu(\hat{G}^r_{\bar{\mu}}) - \nu(\hat{F}_Y) & \text{if } \nu \text{ is a } D_2 \text{ parameter,} \\ \nu(\hat{G}^r_{\bar{m}}) - \nu(\hat{F}_Y) & \text{if } \nu \text{ is a quantile contrast.} \end{cases}$$

for $r \in \{U, L\}$. When ν is linear, we have $\nu(\hat{F}^r) = (1/n)\sum_{i=1}^n \nu(\hat{F}_{Y|X}(\cdot, \hat{S}_\nu^r(W_i, Z_i), Z_i))$ for $r \in \{U, L\}$, where $\hat{S}_\nu^U(w, z) = \hat{H}^{-1}(\hat{F}_W(w-) + \hat{V}_\nu^U(w, z)(\hat{F}_W(w) - \hat{F}_W(w-)))$ and the estimated rank allocator $\hat{V}_\nu^U(w, z)$ is the value of the empirical distribution function of the variables $\hat{\tilde{V}}_{\nu,i} = \nu(\hat{F}_{Y|X}(\cdot|\hat{H}^{-1}(\hat{F}_W(W_i)), Z_i)) - \nu(\hat{F}_{Y|X}(\cdot|\hat{H}^{-1}(\hat{F}_W(W_i-)), Z_i))$, in the group of observations with $W_i = w$, evaluated at $\nu(\hat{F}_{Y|X}(\cdot, \hat{H}^{-1}(\hat{F}_W(w)), z)) - \nu(\hat{F}_{Y|X}(\cdot, \hat{H}^{-1}(\hat{F}_W(w-)), z))$. The function $\hat{S}_\nu^L(w, z)$ is defined analogously, and all other quantities are as given above. For nonlinear functionals ν, estimates of the stochastic dominance bounds G^L and G^U are given by $\hat{G}^r(y) = (1/n)\sum_{i=1}^n \hat{F}_{Y|X}(y|\hat{S}_y^r(W, Z), Z_i)$ for $r \in \{U, L\}$, where $\hat{S}_y^U$ and $\hat{S}_y^L$ can be obtained in the same way as $\hat{S}_\nu^U$ and $\hat{S}_\nu^L$. When ν is a D_2-parameter or a quantile contrast, we compute the respective dispersed and compressed CDFs in Theorem 3 (ii)–(iii) from the estimates $\hat{G}^L$ and $\hat{G}^U$, denoting the result by $\hat{G}^L_{\bar{\mu}}$, $\hat{G}^U_{\bar{\mu}}$, $\hat{G}^L_{\bar{m}}$ and $\hat{G}^U_{\gamma}$, respectively.

In Appendix C, we provide a complete asymptotic theory for our estimators, adapting arguments used in Chernozhukov et al. (2009a) or Rothe (2010). We show that under point identification our estimate of the counterfactual outcome CDF converges to a Gaussian process. Normality of FPPE estimates then follows from the functional delta method. A similar approach is used to establish joint asymptotic normality of the estimated bounds under partial identification. An ordinary bootstrap procedure can be shown to give asymptotically valid approximations to the various Gaussian limit distribution, allowing to circumvent direct estimation of their (often complicated) covariance functions. Then standard methods can be used to construct confidence intervals for the FPPE under point identification. In case of interval-identified parameters, one can use general results in Imbens and Manski (2004) and Stoye (2009).

4.2. Marginal Partial Policy Effects. Under point identification, estimates of the MPPE can be obtained by "plug-in" estimators in a similar fashion as described above. Since these parameters are very similar to the ones discussed in Firpo et al. (2009), we omit a detailed discussion. When W is binary, the MPPE falls into the class of partially identified parameters restricted by intersection bounds, and one can use the methodology proposed by Chernozhukov et al. (2009b). Their approach consist of adding a precision-correction term to a suitable estimate $z \mapsto \hat{g}_\nu(z)$ of the bound-generating function in Theorem 5 before applying the supremum and and infimum operators in order to obtain median unbiased estimates. To be specific, the estimate of the identified set given by $\hat{\mathcal{B}}_W(\nu) = [\hat{\beta}^L_W(\nu), \hat{\beta}^U_W(\nu)]$, where

$$\hat{\beta}^U_W(\nu) = \max_{z \in \hat{\mathcal{Z}}^U} [\hat{g}_\nu(z) - k_{1/2} s(z)] \quad \text{and} \quad \hat{\beta}^L_W(\nu) = \min_{z \in \hat{\mathcal{Z}}^L} [\hat{g}_\nu(z) + k_{1/2} s(z)].$$

Here $\hat{g}_\nu(z)$ is an estimate of the bound generating function $g_\nu(z)$, $s(z)$ is the corresponding pointwise standard error, the critical value $k_{1/2}$ is an estimate of the median of the maximum of the stochastic process $\mathbb{Z}_n(z) := (\hat{g}_\nu(z) - g_\nu(z))/s(z)$, and the sets $\hat{\mathcal{Z}}^U$ and $\hat{\mathcal{Z}}^L$ are both (random) subsets of the support of Z that contain the points where the maximum and minimum is achieved with probability tending to one, respectively. The estimator $\hat{g}_\nu$ can be fully nonparametric or impose parametric restrictions. Its specific form (and thus the choice of s and k_p) depends on the functional ν, and is explicitly described in

Appendix D for the case of the mean and the quantile functional. Chernozhukov et al. (2009b) show that a similar idea can be used to a construct confidence intervals for the parameter of interest, which is valid uniformly with respect to the location of the MPPE within the bounds. We discuss the details in Appendix D.

5. Concluding Remarks

In this paper, we propose a method to evaluate the effect of a counterfactual change in the marginal distribution of a single covariate on the unconditional distribution of an outcome variable of interest. We show that such effects are point identified under general conditions if the covariate affected by the counterfactual change is continuously distributed, but typically only partially identified if its distribution is discrete. For the latter case, we derive informative bounds making use of the available information.

In Appendix E, we present an illustrative empirical application of our methodology, that investigates the role of composition effects in the polarization of the US labor market observed since the mid 1980s (Autor, Katz, and Kearney, 2006). We use our FPPEs to quantify to what extend changes in the unconditional distribution of various observable characteristics of the labor force contributed to the rise in wages at both the top and bottom end of the wage distribution. The results suggest that changes in education and labor market experience only had a minor impact. De-unionization is shown to have contributed to the increase in both overall wage inequality and inequality at the top-end of the wage distribution. However, due to the binary nature of individual union coverage, the effect is only partially identified and thus can only be bounded.

A. Proofs of Theorems

Proof of Lemma 1. We only prove part (i). The proof of (ii) is similar. Using the conditional exogeneity condition in Assumption 1, we find that

$$\begin{aligned}
F_Y^H(y) &= \Pr(m(X^H, \eta) \leq y) \\
&= \int \Pr(m(H^{-1}(u), z, \eta) \leq y | U_1 = u, Z = z) dF_{UZ}(u, z) \\
&= \int \Pr(m(w, z, \eta) \leq y | Z = z) dF_{UZ}(H(w), z) \\
&= \int \Pr(m(w, z, \eta) \leq y | Q_W(U_1) = w, Z = z) dF_{UZ}(H(w), z) \\
&= \int \Pr(m(W, Z, \eta) \leq y | W = w, Z = z) dF_{UZ}(H(w), z) \\
&= \int F_{Y|X}(y, w, z) dF_{UZ}(H(w), z) \\
&= \int F_{Y|X}(y, H^{-1}(u), z) dF_{UZ}(u, z) \\
&= \mathbb{E}(F_{Y|X}(y, H^{-1}(U_1), Z)),
\end{aligned}$$

as claimed. □

Proof of Theorem 1. Under the conditions of the Theorem, there exists a one-to-one relationship between W and U_1 over the range on which H^{-1} is not constant. Hence we have that $F_Y^H(y) = \mathbb{E}(F_{Y|X}(y, H^{-1}(F_W(W)), Z))$. Assumption 1 (ii) ensures that that $F_{Y|X}$ is identified over the area of integration on the right-hand side of the last equation. Hence F_Y^H is identified, and so are of course population parameters of the form $\nu_H = \nu(F_Y^H)$. □

Proof of Lemma 2. The set $\mathcal{F}_Y^H$ of feasible counterfactual outcome distributions is defined as the set of all CDFs F which can be written as $F(y) = \mathbb{E}(F_{Y|X}(y, H^{-1}(\tilde{U}_1), Z))$ for some random variable $\tilde{U}_1 \sim U[0, 1]$ such that $(Q_W(\tilde{U}_1), Z) \stackrel{d}{=} (W, Z)$. Let $\tilde{U}_1$ be any random variable satisfying these two conditions, and let $V = \tilde{U}_1/(F_W(W) - F_W(W-)) - F_W(W-)$. Then V is a rank allocator in the sense of the Lemma, since $\tilde{U}_1 | W = w \sim U[F_W(w-), F_W(w))]$. On the other hand, it is easy to see that $\tilde{F}_W(W, V) \sim U[0, 1]$ and $(Q_W(\tilde{F}_W(W, V)), Z) \stackrel{d}{=} (W, Z)$ for any rank allocator V. In particular, the latter statement follows from the fact that Q_W is constant on the interval $[F_W(w-), F_W(w))]$ for all $w \in \text{supp}(W)$. □

Proof of Theorem 2. The proof for the case that both H and the distribution of W are binary is given in the main text. The proof for the general case is completely analogous and thus omitted. Sharpness of the bounds follows from the fact that by Lemma 2 every valid rank allocator corresponds to a feasible counterfactual outcome distribution, and vice versa. □

Proof of Theorem 3. We first show that G^L and G^U are proper distribution functions, that constitute best possible bounds with respect to first-order stochastic dominance ordering, in the sense that

$$G^U \succeq_1 F \succeq_1 G^L \text{ for all } F \in \mathcal{F}_Y^H, \tag{A.1}$$

and that there do not exist distribution functions $\tilde{G}^L$ and $\tilde{G}^U$ such that $G^U \succ_1 \tilde{G}^U \succ_1 F$ or $F \succ_1 \tilde{G}^L \succ_1 G^L$ for all $F \in \mathcal{F}_Y^H$. From Theorem 2, it follows directly that $G^U(y) \leq F(y) \leq G^L(y)$ for all $F \in \mathcal{F}_Y^H$ and all $y \in \mathbb{R}$, since the functional ν with $\nu(F) = F(y)$ is linear. This proves the claim in (A.1).

Next, we show that G^L and G^U are CDFs. It is immediate by construction that both functions are right-continuous and tend to zero and one as the point of evaluation tends to $\pm\infty$. It thus remains to be shown that they are nondecreasing. To see this, note that by the sharpness result in Theorem 2, for any $\bar{y} \in \mathbb{R}$ there exists a feasible counterfactual outcome distribution $\bar{F}_{\bar{y}} \in \mathcal{F}_Y^H$ such that $\bar{F}_{\bar{y}}(y) = G^U(y)$ for $y = \bar{y}$. Now suppose G^U was not everywhere nondecreasing, i.e. $G^U(y') < G^U(y)$ for some $y' > y$. This would imply that $G^U(y) > \bar{F}_{y'}(y') \geq \bar{F}_{y'}(y)$ since $\bar{F}_{y'}$ is a proper CDF, which violates the fact that $G^U(y) \leq F(y)$ for all $F \in \mathcal{F}_Y^H$ and all $y \in \mathbb{R}$. Hence G^U must be nondecreasing. An analogous argument applies to G^L.

Finally, we show that G^L and G^U are best possible bounds with respect to the (partial) ordering induced by stochastic dominance. Suppose there exists a function $\tilde{G}^U$ such that $\tilde{G}^U(y) \geq G^U(y)$ for all $y \in \mathbb{R}$, and $\tilde{G}^U(\bar{y}) > G^U(\bar{y})$ for some $\bar{y} \in \mathbb{R}$. Then by Theorem 2, there exists a feasible counterfactual outcome distribution $F \in \mathcal{F}_Y^H$ such that $G^U(\bar{y}) = F(\bar{y})$. Hence it cannot be the case that $\tilde{G}^U(y) \leq F(y)$ for all $F \in \mathcal{F}_Y^H$ and all $y \in \mathbb{R}$. An analogous argument applies to G^L.

With these arguments, the statement of the Theorem follows directly from Theorem 1 and 2 in Stoye (2010). □

Proof of Theorem 4. This follows from application of the chain rule and the representation of F_Y^H in Lemma 2(ii). $\square$

Proof of Theorem 5. First, note that whenever the distribution of W is not degenerate, i.e. $F_W(0) \in (0,1)$, we have that $C_1(F_W(0), \cdot) \in \mathcal{S}$, where $\mathcal{S}$ is the set of all multivariate distribution functions with support $R_{\mathcal{Z}} = \{(F_{Z_1}(z_1), \dots, F_{Z_{d'}}(z_{d'})) : z \in \mathcal{Z}\}$, where $\mathcal{Z}$ denotes the support of X. For the case that the dimension of (W, Z) is equal to two, i.e. $d' = 1$, this follows from Theorem 2.2.7 in Nelsen (2006). The extension of his result to the general multivariate case is immediate.

Next, let $\mathcal{T} = \{T : T(z) = s(F_{Z_1}(z_1), \dots, F_{Z_{d'}}(z_{d'})), s \in \mathcal{S}\}$. Note that it follows from the properties of $\mathcal{S}$ that $\mathcal{T}$ is the set of all distribution functions with support $\mathcal{Z}$. It then follows directly that

$$\inf_{z \in \mathcal{Z}} g_\nu(z) \leq \sup_{T \in \mathcal{T}} \int g_\nu(z) dT(z) \leq \sup_{z \in \mathcal{Z}} g_\nu(z).$$

Since $\mathcal{T}$ is the set of *all* distribution functions with support $\mathcal{Z}$, these bounds are sharp. $\square$

B. Additional Definitions

In this section, we give the precise definition of three important concepts, which are omitted from the main body of the paper for brevity: the distributional features covered by Theorem 3, the notion of a compressed and dispersed distribution necessary to state the bounds in Theorem 3(ii)–(iii), and that of Hadamard differentiability.

Definition 2 (Distributional Features (Stoye, 2010)). *Consider a functional $\nu : \mathcal{F} \to \mathbb{R}$.*

i) ν is a D_1-parameter if it increases with first-order stochastic dominance, i.e $F \succeq_1 G$ implies that $\nu(F) \geq \nu(G)$.

ii) ν is a D_2-parameter if it increases with second-order stochastic dominance for any two distributions that have the same mean, i.e. $\mu(F) = \mu(G)$ and $F \succeq_2 G$ implies $\nu(F) \geq \nu(G)$.

iii) ν is an (α, β)-quantile contrast if $\nu(F) = g(F^{-1}(\alpha), F^{-1}(\beta))$ for $\alpha \leq \beta$ and a known function $g : \mathbb{R}^2 \to \mathbb{R}$ that is non-increasing in the first and nondecreasing in the second argument.

Definition 3 (Compressed and Dispersed Distributions (Stoye, 2010)). *The distribution function $F^C(\cdot) = F^C(\cdot|a, F^U, F^L)$ is called compressed relative to two other distribution F^L and F^U with threshold value $a \in \mathbb{R}$ if*

$$F^C(y) = \begin{cases} F^U(y), & y < a \\ F^L(y), & y \geq a. \end{cases}$$

The distribution function $F^D(\cdot) = F^D(\cdot|a, F^U, F^L)$ is called dispersed relative to two other distributions F^L and F^U with threshold value $a \in [0, 1]$ if

$$F^D(y) = \begin{cases} F^L(y), & y < Q^L(a) \\ a, & Q^L(a) \leq y < Q^U(a) \\ F^U(y), & y \geq Q^U(a). \end{cases}$$

Definition 4 (Hadamard Differentiability (Van der Vaart, 2000)). *The functional $\nu : \mathcal{F} \to \mathbb{R}$ is called Hadamard differentiable at F if there exists a continuous map $\nu'_F : \mathcal{F} \to \mathbb{R}$ such that*

$$\left| \frac{\nu(F + th_t) - \nu(F)}{t} - \nu'_F(h) \right| \to 0$$

as $t \to 0$, for every $h_t \to h$ such that $F + th_t$ is contained in the domain of ν for all values of t sufficiently close to zero.

C. Asymptotic Theory for Fixed Partial Policy Effect Estimators

In this section, we investigate the asymptotic properties of the estimators proposed in Section 4.1. For our asymptotic analysis, we adapt arguments used in Chernozhukov et al. (2009a) or Rothe (2010) for estimators of distributional policy effects corresponding to changes in the entire covariate distribution. We show that under point identification our estimate of the counterfactual outcome CDF converges to a Gaussian process. Normality of sufficiently smooth population parameters then follows from the functional delta method. A similar approach is used to establish joint asymptotic normality for the estimates of the upper and lower bounds of the various identified sets. Such results can be used to construct asymptotically valid confidence regions for the objects of interest.

In order to account for both nonparametric and parametric estimates of the conditional CDF $F_{Y|X}$, we conduct our analysis under general "high-level" assumptions, that can be verified for a wide range of estimation procedures under standard regularity conditions. The assumptions are stated in such a way that the respective theorems follow by straightforward arguments, using the Donsker Theorem, the Functional Delta Method and the Continuous Mapping Theorem (Van der Vaart, 2000). We thus omit all proofs.

A word on notation: we denote the support of Y, W, Z, X and W^* by $\mathcal{Y}$, $\mathcal{W}$, $\mathcal{Z}$, $\mathcal{X}$ and $\mathcal{W}^*$, respectively. The space $\ell^\infty(A)$ is the space of all uniformly bounded functions mapping from A to $\mathbb{R}$, equipped with the metric induced by the supremum norm. We also write "$\xrightarrow{d}$" to denote convergence in distribution of a sequence of random variables, and "$\Rightarrow$" to denote weak convergence of a sequence of random functions.

C.1. Estimators under Point Identification. To avoid notational complications, we assume that both F_W and H are continuous and strictly increasing. We derive our results under the following high-level conditions.

Assumption 1. *(i) Let* $\mathbf{G}_Y^H(y) = \int \mathbf{F}_{Y|X}(y, H^{-1}(F_W(w)), z) dF_X(w,z)$*, where we write* $\mathbf{F}_{Y|X}(y,w,z) = \sqrt{n}(\hat{F}_{Y|X}(y,w,z) - F_{Y|X}(y,w,z))$*, and define the processes* $\mathbf{F}_X(w,z) = \sqrt{n}(\hat{F}_X(w,z) - F_X(w,z))$*,* $\mathbf{F}_W(w) = \sqrt{n}(\hat{F}_W(w) - F_W(w))$*, and* $\mathbf{Q}_W^*(\tau) = \sqrt{n}(\hat{H}^{-1}(\tau) - H^{-1}(\tau))$*. Then*

$$(\mathbf{G}_Y^H, \mathbf{F}_X, \mathbf{F}_W, \mathbf{Q}_W^*) \Rightarrow (\mathbb{G}_Y^H, \mathbb{F}_X, \mathbb{F}_W, \sqrt{\lambda}\mathbb{Q}_W^*)$$

in the space $\ell^\infty(\mathcal{Y}) \times \ell^\infty(\mathcal{X}) \times \ell^\infty(\mathcal{W}) \times \ell^\infty([0,1])$*, where the right hand side is a mean zero Gaussian process. (ii) The function class* $\{(w,z) \mapsto F_{Y|X}(y, H^{-1}(F_W(w)), z), y \in \mathbb{R}\}$ *is* F_X*-Donsker. (iii) The partial derivative* $\partial_w F_{Y|X}(y,w,z)$ *exists for all* $(y,w,x) \in \mathcal{Y} \times \mathcal{W} \times \mathcal{X}$ *and is uniformly bounded. (iv)* $\sup_{(y,w,x)} |\partial_w \hat{F}_{Y|X}(y,w,z) - \partial_w F_{Y|X}(y,w,z)| = o_p(1)$*.*

The first part of Assumption 1 can e.g. be verified using results in Chernozhukov et al. (2009a), who establish convergence in distribution of $\hat{F}_{Y|X}$ to a Gaussian process for a variety of different CDF estimators involving certain parametric restrictions. The condition then follows directly from the continuous mapping theorem, and the fact that the 2nd–4th component of the process are just empirical CDFs and quantile functions in our case. Assumption 1(i) can also be verified by direct arguments if $\hat{F}_{Y|X}$ is a nonparametric estimator converging at a rate slower than $n^{-1/2}$ to a limit process which is not tight. For example, Rothe (2010) proves such a condition for a kernel-based estimator of the conditional CDF, using the theory of U-processes. We conjecture that the assumption could also be verified for other nonparametric estimators, such as e.g. those based on sieves or orthogonal series. Assumption 1(ii) is a weak regularity condition fulfilled by various classes of functions (e.g. Van der Vaart, 2000, Chapter 19). Finally, Assumption 1(iii)-(iv) are weak smoothness conditions on the estimated CDF and its population counterpart.

Theorem 6. *Suppose that Assumption 2 holds. Then the process* $\mathbf{F}_Y^H = \sqrt{n}(\hat{F}_Y^H - F_Y^H)$ *converges weakly to the following mean zero Gaussian process:*

$$\mathbf{F}_Y^H \Rightarrow \mathbb{F}_{Y,A}^H + \mathbb{F}_{Y,B}^H + \mathbb{F}_{Y,C}^H =: \mathbb{F}_Y^H$$

where

$$\begin{aligned}
\mathbb{F}_{Y,A}^H(y) &= \int \mathbb{F}_{Y|X}(y, H^{-1}(F_W(w)), z) dF_X(w,z), \\
\mathbb{F}_{Y,B}^H(y) &= \int F_{Y|X}(y, H^{-1}(F_W(w)), z) d\mathbb{F}_X(w,z), \\
\mathbb{F}_{Y,C}^H(y) &= \int \partial_w F_{Y|X}(y, H^{-1}(F_W(u)), z) \\
&\qquad \times \left(\partial_\tau H^{-1}(F_W(u))\mathbb{F}_W(u) + \sqrt{\lambda}\mathbb{Q}_W^*(F_W(u))\right) dF_X(u,z),
\end{aligned}$$

and the convergence is in $\ell^\infty(\mathcal{Y})$.

Corollary 1. *Suppose that the conditions of Theorem 6 hold, and that the functional $\nu : \mathcal{F} \to \mathcal{G}$ is Hadamard differentiable at F_Y^H with derivative ν'. Then*

$$\sqrt{n}(\hat{\alpha}_W(\nu) - \alpha_W(\nu)) \xrightarrow{d} \nu'(\mathbb{F}_Y^H - \mathbb{F}_Y),$$

where the right-hand side is a Normal distribution with mean zero.

C.2. Bounds on Linear Functionals. In this subsection, we assume that both W and W^* are discretely distributed, ruling out the mixed discrete-continuous case to avoid notational complications. We make the following assumptions.

Assumption 2. *(i) Using the notation introduced in Assumption 1, and defining* $\mathbf{G}^H_{Y,s} = \int \mathbb{F}_{Y|X}(y, S^s_\nu(w,z), z) dF_X(w,z)$ *for* $s \in \{U, L\}$*, have that*

$$(\mathbf{G}^H_{Y,U}, \mathbf{G}^H_{Y,L}, \mathbf{F}_X) \Rightarrow (\mathbb{G}^H_{Y,U}, \mathbb{G}^H_{Y,L}, \mathbb{F}_X)$$

in the space $\ell^\infty(\mathcal{Y}) \times \ell^\infty(\mathcal{Y}) \times \ell^\infty(\mathcal{X})$*, and the right hand side is given by a mean zero Gaussian process. (ii) The function class* $\{(w,z) \mapsto F_{Y|X}(y, S^s_\nu(w,z), z), y \in \mathbb{R}, s \in \{U, L\}\}$ *is* F_X*-Donsker. (iii)* $\Pr(\hat{S}^s_\nu(W,Z) = S^s_\nu(W,Z)) \to 1$ *as* $n \to \infty$ *for* $s \in \{U, L\}$*.*

Assumption 2 (i)–(ii) is similar to Assumption 1 (i)–(ii) but otherwise analogous, and hence the same comments apply. Assumption 2 (iii) naturally holds in our setting, since W^* is discrete and the rank allocator variables are estimated consistently. To see this, note that for the empirical quantile function of a discrete random variable it holds that $\Pr(\hat{Q}^*_W(\tau) = Q^*_W(\tau)) \to 1$ as $n \to \infty$ for all τ except those in a set of measure zero.

Theorem 7. *Suppose that Assumption 2 holds, and that* ν *is linear. Then the terms* $\mathbf{N}^s = \sqrt{n}(\hat{\alpha}^s_W(\nu) - \alpha^s_W(\nu))$*,* $s \in \{U, L\}$*, jointly converge in distribution to a Normal distribution with mean zero:*

$$\sqrt{n}(\mathbf{N}^L, \mathbf{N}^U) \xrightarrow{d} (\nu(\mathbb{F}^L), \nu(\mathbb{F}^U))$$

where $\mathbb{F}^s = \mathbb{F}^s_A + \mathbb{F}^s_B$*,* $s \in \{U, L\}$ *is a mean zero Gaussian process with*

$$\mathbb{F}^s_A(y) = \int \mathbb{F}_{Y|X}(y, S^s_\nu(w,z), z) dF_X(w,z),$$
$$\mathbb{F}^s_B(y) = \int F_{Y|X}(y, S^s_\nu(w,z), z) d\mathbb{F}_X(w,z),$$

and the convergence is in $\ell^\infty(\mathcal{Y})$*.*

C.3. Bounds on Smooth Functionals. To obtain asymptotic properties for the estimated boundaries of the identified set based on the result in Theorem 3, we maintain the assumption that W and W^* are discretely distributed. We also maintain the first part of Assumption 2, modifying the remainder as follows.

Assumption 3. *(i) Using the notation introduced in Assumption 1, and defining* $\mathbf{G}^H_{Y,s} = \int \mathbb{F}_{Y|X}(y, S^s_y(w,z), z) dF_X(w,z)$ *for* $s \in \{U, L\}$*, have that*

$$(\mathbf{G}^H_{Y,U}, \mathbf{G}^H_{Y,L}, \mathbf{F}_X) \Rightarrow (\mathbb{G}^H_{Y,U}, \mathbb{G}^H_{Y,L}, \mathbb{F}_X)$$

in the space $\ell^\infty(\mathcal{Y}) \times \ell^\infty(\mathcal{Y}) \times \ell^\infty(\mathcal{X})$, and the right hand side is given by a mean zero Gaussian process. (ii) The function class $\{(w,z) \mapsto F_{Y|X}(y, S_y^s(w,z), z), y \in \mathbb{R}, s \in \{U, L\}\}$ is F_X-Donsker. (iii) $\Pr(\hat{S}_y^s(W,Z) = S_y^s(W,Z)$ for all $y \in \mathcal{Y}) \to 1$ as $n \to \infty$ for $s \in \{U, L\}$.

Assumption 3 only constitutes a minor modification of Assumption 2, adjusting for the fact that the rank allocator variable used to construct the upper and lower bounding functions varies with the point of evaluation.

Theorem 8. *Suppose that Assumption 3 holds. Then the processes $\mathbf{G}^s = \sqrt{n}(\hat{G}^s - G^s)$ converge weakly to mean zero Gaussian processes, jointly over $s \in \{U, L\}$:*

$$\mathbf{G}^s \Rightarrow \mathbb{G}_A^s + \mathbb{G}_B^s =: \mathbb{G}^s, \quad s \in \{U, L\}$$

where

$$\mathbb{G}_A^s(y) = \int \mathbb{F}_{Y|X}(y, S_y^s(w,z), z) dF_X(w,z)$$
$$\mathbb{G}_B^s(y) = \int F_{Y|X}(y, S_y^s(w,z), z) d\mathbb{F}_X(w,z)$$

and the convergence is in $\ell^\infty(\mathcal{Y})$.

To use this result to derive asymptotic properties of the estimated boundaries of the identified set, we introduce the following assumption concerning the smoothness of the population parameter of interest with respect to the underlying distribution.

Assumption 4. *(i) The functional the functional $\nu : \mathcal{F} \to \mathbb{R}$ is Hadamard differentiable at F_Y^H with derivative ν'. (ii) Let $a_C^* = a_C^*(\bar{\mu}, G^U, G^L)$ and $a_D^* = a_D^*(\bar{\mu}, G^U, G^L)$ be threshold values yielding a compressed or dispersed distribution relative to G^U and G^L with mean $\bar{\mu}$, respectively. That is,*

$$\int y dF^C(y|a_C^*, G^U, G^L) = \bar{\mu} \text{ and } \int y dF^D(y|a_D^*, G^U, G^L) = \bar{\mu}.$$

Then the map $(F_1, F_2) \mapsto T_{\bar{\mu}}^s(F_1, F_2) = \nu(F^s(\cdot|a_s^*(\bar{\mu}, F_1, F_2), F_1, F_2))$ is Hadamard differentiable at (G^L, G^U) with derivative $T_{\bar{\mu}}^{s\prime}(F_1, F_2)$ for $s \in \{C, D\}$.

Assumption 4(i) can be verified for most common distributional features of interest under standard regularity condition, e.g. moments, quantiles, or the Gini coefficient

(Rothe, 2010). Assumption 4(ii) is necessary to analyze the estimated bounds on D_2-parameters, since e.g. the mapping that transforms two CDFs into a compressed distribution with a particular mean is not Hadamard differentiable due to the discontinuity at the threshold value. A sufficient condition for Assumption 4(ii) is that Assumption 4(i) holds, that G^U and G^L are continuous, and that for a compressed or dispersed distribution F the parameter $\nu(F)$ does not depend on the value of F at the threshold value. Using the notation that $S_C(y|a, f_1, f_2) = f_1(y)\mathbb{I}\{y < a\} + f_2(y)\mathbb{I}\{y \geq a\}$ and $S_D(y|a, f_1, f_2) = f_2(y)\mathbb{I}\{y < f_2^{-1}(a)\} + a\mathbb{I}\{f_2^{-1}(a) \leq y < f_1^{-1}(a)\} + f_1(y)\mathbb{I}\{y \geq f_1^{-1}(a)\}$, we can now state the final corollary, which follows directly Theorem 8 and the Functional Delta Method.

Corollary 2. *Suppose that Assumption 4(i) and the conditions of Theorem 8 hold. Then the terms $\mathbf{N}^s = \sqrt{n}(\hat{\alpha}_W^s(\nu) - \alpha_W^s(\nu))$, $s \in \{U, L\}$, jointly converge in distribution to a Normal distribution with mean zero:*

i) If ν is a D_1-parameter, then

$$\sqrt{n}(\mathbf{N}^L, \mathbf{N}^U) \xrightarrow{d} (\nu'(\mathbb{G}^L - \mathbb{F}_Y), \nu'(\mathbb{G}^U - \mathbb{F}_Y)).$$

ii) If ν is a quantile contrast, then

$$\sqrt{n}(\mathbf{N}^L, \mathbf{N}^U) \xrightarrow{d} (\nu'(S_C(\cdot|\bar{m}, \mathbb{G}^U, \mathbb{G}^L) - \mathbb{F}_Y), \nu'(S_D(\cdot|\gamma, \mathbb{G}^U, \mathbb{G}^L) - \mathbb{F}_Y)).$$

iii) If ν is a D_2-parameter and Assumption 4(ii) holds, then

$$\sqrt{n}(\mathbf{N}^L, \mathbf{N}^U) \xrightarrow{d} (T_{\bar{\mu}}^{C\prime}(\mathbb{G}^U, \mathbb{G}^L) - \nu'(\mathbb{F}_Y), T_{\bar{\mu}}^{D\prime}(\mathbb{G}^U, \mathbb{G}^L) - \nu'(\mathbb{F}_Y)).$$

In each case, the right-hand side is a bivariate Normal distribution with mean zero.

C.4. Inference. Our results in the previous subsections imply that under general conditions our objects of interest are asymptotically normal. Under point identification, this insight can be used to construct confidence intervals for identified features in the usual fashion. Under partial identification, our results imply that confidence regions for the various identified sets can be formed by computing one-sided confidence regions for its upper and lower boundaries. This can be done in the same way as in the point identified case. If the interest is in obtaining a confidence region for the population parameter of interest, as opposed to the identified set, this can be accomplished by using the general results on inference for interval-identified parameters in Imbens and Manski (2004) and Stoye (2009).

In both cases, the major complication is that the covariance function of the limiting Gaussian distributions can be quite complicated to compute directly. However, it follows from results in Chernozhukov et al. (2009a) or Rothe (2010) that under both point identification and partial identification an ordinary bootstrap procedure can be used to approximate the various limiting distributions of the previous subsection in finite samples. This result can be shown to hold for parametric and nonparametric estimation procedures of the conditional CDF $F_{Y|X}$, and thus provides a straightforward and tractable way to conduct inference in empirical applications.

We also remark that our results in the previous subsections immediately generalize to function-valued population parameters, allowing researchers conduct uniform inference on the counterfactual outcome distribution under essentially the same conditions. That is, it is not only possible to compute confidence intervals for a real-valued population parameter, but also to compute uniform confidence regions for function valued parameters, such as the CDF itself, or the corresponding quantile process. This is an important feature of our results, as it allows applied researchers to test hypotheses that cannot be adequately addressed by considering only a fixed number of isolated points. An example would be an hypothesis such as "The change in the marginal distribution of W to that of W^* did not affect the outcome distribution".

D. Asymptotic Theory for Marginal Partial Policy Effect Estimators

In this section, we describe some of the details about how to construct estimates of our Marginal Partial Policy Effects, and how to derive their theoretical properties. We focus on the partially identified case of a binary covariate, since under point identification such results follow from standard arguments. In particular, when W is continuous a nonparametric sample analogue estimator based on the identification result in Theorem 4 would be very similar to an average derivative estimator, which can e.g. be analyzed using results in Newey (1994). See also Firpo et al. (2009) for a similar analysis.

As one can see from Theorem 5, the identified set of the MPPE in case of a binary covariate is restricted by the extrema of the "bound generating function" $z \mapsto g_\nu(z)$. The problem thus 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 et al. (2009b), henceforth abbreviated CLR. Our paper does not contain new insights on this issue. In the following, we simply show how to apply their main results to our context.

The basic idea of CLR is to add suitable precision-correction terms to a standard estimate of the bound generating function g_ν before applying the maximum or minimum operator. To explain this in detail, we first have to introduce some notation.[6] For any $p \in (0, 1)$, we define

$$\hat{\beta}^U_W(\nu; p) = \max_{z \in \hat{\mathcal{Z}}^U} [\hat{g}_\nu(z) - k_p s(z)] \quad \text{and} \quad \hat{\beta}^L_W(\nu; p) = \min_{z \in \hat{\mathcal{Z}}^L} [\hat{g}_\nu(z) + k_p s(z)].$$

Here $\hat{g}_\nu(x)$ is an estimate of the bound generating function $g_\nu(x)$, which can be fully nonparametric or impose parametric restrictions, $s(x)$ is the corresponding standard error, the critical value k_p is an estimate of the p-quantile of the maximum of the stochastic process

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

and the sets $\hat{\mathcal{Z}}^U$ and $\hat{\mathcal{Z}}^L$ are both (random) subsets of the support of Z that contain the points where the maximum and minimum is achieved with probability tending to one, respectively. Specifically, CLR recommend to set

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

[6]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

The specific choices of $\hat{g}_\nu$, s and k_p (and thus also those of $\hat{\mathcal{Z}}^U$ and $\hat{\mathcal{Z}}^L$) depend on the Hadamard derivative of the functional ν, and are explicitly described below for the case of the mean and the quantile functional. Finally, define the interval $\hat{\mathcal{B}}_W(\nu, p)$ as

$$\hat{\mathcal{B}}_W(\nu, p) = [\hat{\beta}^L_W(\nu; p), \hat{\beta}^U_W(\nu; p)].$$

With this notation, the estimate of the identified set of the FPPE is then given by $\hat{\mathcal{B}}_W(\nu; 1/2)$. In particular, using the choices described below, Theorem 1 in CLR implies that $\hat{\beta}^U_W(\nu; 1/2)$ is a consistent and asymptotically median unbiased estimate of the upper bound $\beta^U_W(\nu)$ of the identified set, in the sense that

$$\Pr(\beta^U_W(\nu) \leq \hat{\beta}^U_W(\nu; 1/2)) = 1/2 + o(1).$$

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 $\Delta_n = \hat{\beta}^U_W(\nu; 1/2) - \hat{\beta}^L_W(\nu; 1/2)$, and $\hat{p}_n = \Phi(\tau_n \Delta_n^+)c$, where $\Phi(\cdot)$ is the standard normal CDF and $\tau_n = \log(n)/\max[\hat{\beta}^U_W(\nu; 3/4) - \hat{\beta}^U_W(\nu; 1/4), \hat{\beta}^L_W(\nu; 3/4) - \hat{\beta}^L_W(\nu; 1/4)]$. Then $\hat{\mathcal{B}}_W(\nu; \hat{p}_n)$ provides an asymptotic $1 - c$ confidence interval for the parameter of interest, such that

$$\inf_{\beta \in \mathcal{B}_W(\nu)} \Pr(\beta \in \hat{\mathcal{B}}_W(\nu; \hat{p}_n)) \geq 1 - c + o(1).$$

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

We now illustrate the choice of $\hat{g}_\nu$ and s for the case that the functional ν maps a CDF into either its mean or one of its quantiles. Given these choices, CLR describe how to obtain the critical value k_p via simulation methods or an analytical formula. We refer to their Appendix C for a detailed description of the practical implementation.

We start by consider the case where the functional of interest is the mean functional $\mu : F \mapsto \int y dF(y)$. Since μ is linear, it is also Hadamard differentiable, with the derivative being equal to μ itself. It follows that the function g_μ is given by

$$g_\mu(z) = \mathbb{E}(Y|W = 1, Z = z) - \mathbb{E}(Y|W = 0, Z = z).$$

This is simply the difference between two conditional expectations, which, depending on the application, can estimated by a variety of parametric, semiparametric and nonparametric methods. The calculation of standard errors is also straightforward in this case.

We now consider the case where the functional of interest is the quantile functional $\nu_{Q,\tau} : F \mapsto \inf\{y \in \mathbb{R} : F(y) \geq \tau\} := Q(\tau)$, which maps a CDF into the corresponding τ-quantile. If F_Y is continuously differentiable in some open neighborhood of $Q_Y(\tau)$, and its derivative f_Y is strictly positive, it follows from Lemma 21.4 in Van der Vaart (2000) that $\nu_{Q,\tau}$ is Hadamard differentiable with derivative

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

In this case, the bound generating function g_ν simplifies to

$$g_\nu(z) = -\frac{F_{Y|X}(Q_Y(\tau)|1,z) - F_{Y|X}(Q_Y(\tau)|0,z)}{f_Y(Q_Y(\tau))},$$

which can be estimated by substituting sample analogues for all unknown quantities:

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

Here $\hat{Q}_Y$ is the empirical sample quantile function of the observed outcomes, and $\hat{f}_Y$ is a nonparametric kernel density estimator 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 $h \to 0$ we have $nh \to \infty$. Finally, the conditional distribution function $F_{Y|X}$ can be estimated by either of the parametric methods discussed in Chernozhukov et al. (2009a), e.g. by first estimating a linear quantile regression model $Q_{Y|X}(\tau, x) = x'\beta(\tau)$, and then inverting the corresponding conditional quantile quantile function, or by a fully nonparametric CDF estimator, e.g. a kernel estimator as in Rothe (2010).

The construction of appropriate standard errors depends on the choice of conditional CDF estimator. When $F_{Y|X}$ is estimated by fully nonparametric methods, its rate of convergence is typically going to be slower than that of either $\hat{Q}_Y$ and $\hat{f}_Y$, and hence the sampling variation in the latter two quantities can be ignored. When $F_{Y|X}$ is estimated by parametric methods, such as the ones described in Chernozhukov et al. (2009a), it converges at the same $\sqrt{n}$-rate as the quantile function $\hat{Q}_Y$, which is faster than the one-dimensional nonparametric rate of the density estimator $\hat{f}_Y$. From an asymptotic point of view, it would thus be valid to compute standard errors that only account for the sampling variation in $\hat{f}_Y$. In practice, it can still be advisable to include "higher-order" components into the standard errors, which account for the uncertainty in $\hat{Q}_Y$ and $F_{Y|X}$. Those can be obtained via the usual Delta method, and shown to satisfy the conditions in CLR.

E. Empirical Illustration

In this section, we illustrate the application of our methodology, in particular the estimation of FPPEs, by analysing changes in the distribution of male wages in United States from 1985 to 2005. There is now extensive evidence that wage inequality has been rising over this period particularly in top end of the wage distribution, but slightly decreased in the bottom end (e.g. Autor et al., 2006; Lemieux, 2008). Figure 1 illustrates this trend, showing that the change in (log real) wages at each quantile of the wage distribution follows a U-shape. Various explanations have been put forward for this so-called polarization of the labor market, including reduced returns to skilled but "routine" occupations due to the introduction of new information technology (e.g. Autor et al., 2006; Goos and Manning, 2007), and simple mechanical effects due to changes in the composition of the labor force (e.g. Lemieux, 2006, 2008).

In this application, we consider the workforce composition effect as a partial explanation of the polarization of the labor market in greater detail. During the period covered by the data, unionization among US male workers fell from 27% to 15%. On average, the workforce also became older and better educated. Our aim is to quantify to what extend these changes contributed to the overall change of the wage distribution, both individually and as a whole.

Our analysis employs the same dataset as in Firpo et al. (2010)[7], extracted from

[7] I would like to thank Nicole Fortin for making the data from Firpo et al. (2010) available.

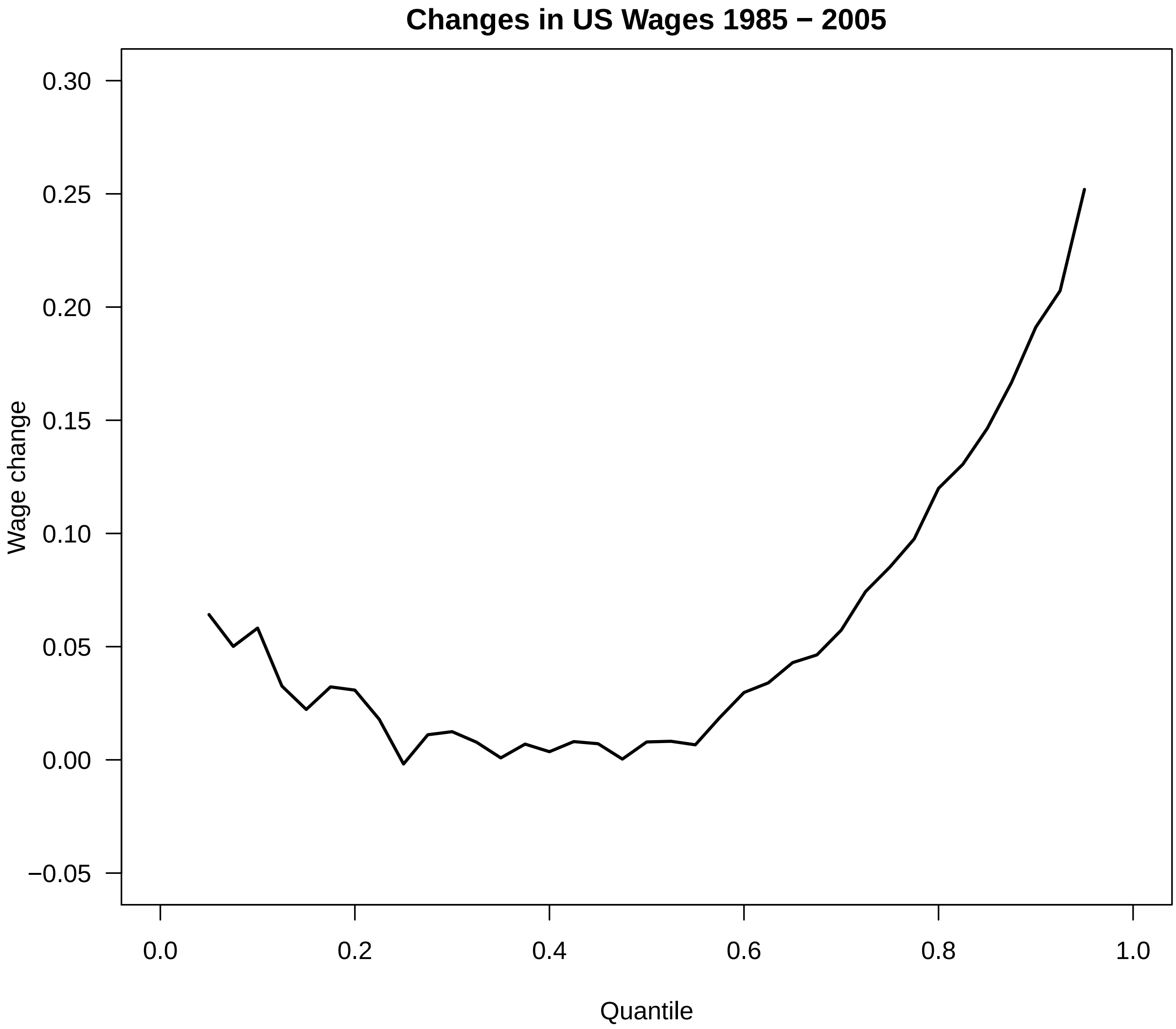

Figure 1: Differences between τ-quantiles of the distribution of US males' log real wages in the 1985 and 2005 for $\tau \in (.05, .95)$.

the the 1983—1985 and 2003–2005 Outgoing Rotation Group (ORG) supplements of the Current Population Survey (CPS). It contains information on 232,784 and 170,693 males, respectively, that were employed in the relevant periods. The data from 1983—1985 play the role of (Y, X), whereas data from 2003–2005 will be used to estimate the direction H of the counterfactual change. The outcome variable Y is the natural logarithm of the hourly wage in 1985 dollars. The covariates X include a dummy for union coverage, years of education, years of potential labor market experience, and dummies for race and marital status, and part-time status. Following common practice, we weigh the observations by

Table 1: Descriptive Statistics

	1985		2005	
	Mean	Std. Dev.	Mean	Std. Dev.
Log Wage	1.785	0.524	1.849	0.583
Education	12.871	2.906	13.420	2.787
Experience	17.199	12.308	19.473	11.486
Married	0.670	0.470	0.620	0.486
Nonwhite	0.113	0.318	0.130	0.336
Union Covered	0.268	0.443	0.153	0.361
Part Time	0.089	0.286	0.093	0.290

Note: Observations are weighted by the product of CPS sample weights and the number of hours worked.

the product of the CPS sampling weights and the hours worked to obtain a representative sample of the total hours worked in the economy. Some descriptive statistics are provided in Table 1.

Estimation is carried out using the procedures described in Section 4.1 of the paper. We estimate the conditional CDF of Y given X by a flexible parametric approach due to Foresi and Peracchi (1995), modeling the conditional probability of the event $(Y \leq y)$ separately for each $y \in \mathbb{R}$ via a logistic regression. That is, we set $F_{Y|X}(y, x) = \Lambda(t(x)'\beta(y))$, where Λ is the Logistic distribution function, $t(\cdot)$ is a known transformation used to generate quadratic, cubic, or interaction terms, and $\beta(y)$ is a finite dimensional parameter indexed by $y \in \mathbb{R}$, that can be estimated by maximum likelihood. This estimator is straightforward to implement and numerically stable. In addition to the covariates mentioned above, we use quadratic terms in education and experience and a full set of interaction terms to estimate $F_{Y|X}$ (37 parameters in total). We estimate the (identified set of the) fixed partial distributional policy effect $\alpha_W(\nu)$ for various functionals ν, with the role of W and W^* being taken by eduction, experience and union coverage, respectively. Results for other covariates are omitted for brevity. We treat the former two quantities as continuously distributed, and derive bounds only for the effect of union coverage. We also estimate the effect of a change in the entire distribution of the explanatory variables (from that in 1983–1985 to that in 2003–2005) using the method in

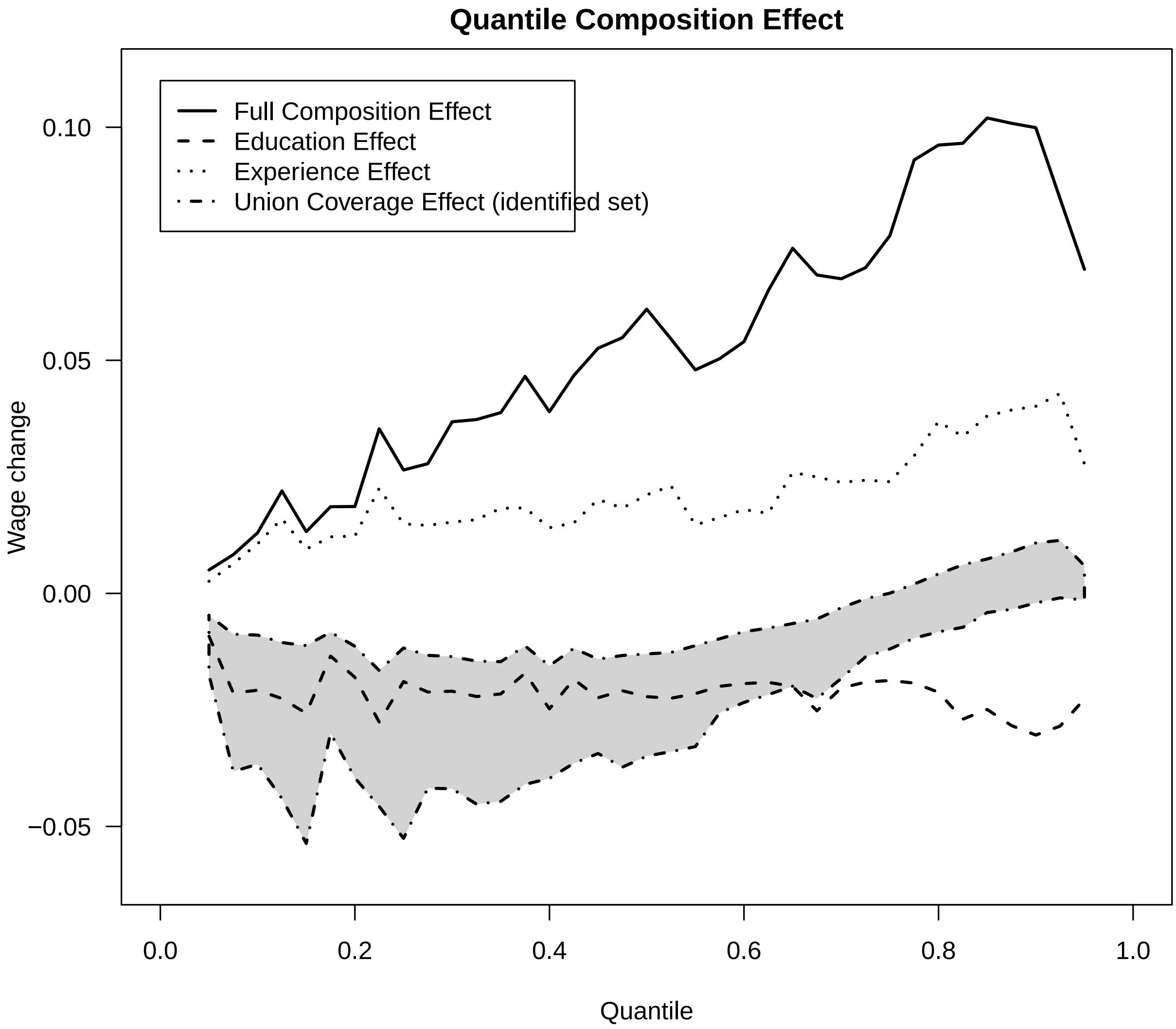

Figure 2: Relative change in τ-quantile of US male wages from 1985–2005 for $\tau \in (.1, .9)$: full distributional policy effect calculated using the method in Chernozhukov et al. (2009a) (bold line); FPPE of changes in education (dashed line); FPPE of changes in experience (dotted line); identified set of FPPE of change in unionization (shaded area).

Chernozhukov et al. (2009a). Results are given in Figure 2– 3 and Table 2– 3. Due to the large size of the data set, those estimates have virtually no relevant sampling variation. We thus omit standard errors, t-statistics and the like.

Figure 2 shows estimates of the full distributional policy effect (as defined in Rothe, 2010) and of the FPPE for various quantiles of the (log real) wage distribution. The full effect can be seen to have contributed to the increase in wage inequality. It has

Table 2: Decomposition Results: Location Measures

	Mean	Q10	Q25	Q50	Q75	Q90
Total Change	0.064	0.058	-0.002	0.008	0.085	0.191
Full Policy Effect	0.058	0.013	0.026	0.061	0.077	0.100
Partial Policy Effect:						
Education	-0.015	-0.021	-0.019	-0.022	-0.019	-0.030
Experience	0.026	0.011	0.015	0.021	0.024	0.040
Union Coverage						
Upper bound	-0.001	-0.009	-0.016	-0.024	0.000	0.027
Lower bound	-0.035	-0.038	-0.074	-0.063	-0.026	-0.005

Note: Sampling variation of estimates is negligible due to large sample size.

positive impact on each quantile, with the magnitude of the effect gradually increasing with the quantile under consideration. Accordingly, Table 3 shows that the full distributional policy effect accounts for about two thirds of the increase in the "90-10 gap" (the difference between the 90% and the 10% quantile). However, it does not explain the U-shaped pattern in Figure 1, which must thus be driven by changes in the structural wage functions.

We now consider the FPPEs of individual covariates. From Figure 2, we see that changes in education alone would have both led to a left shift and a compression of the (real log) wage distribution. Accordingly, Table 2–3 show a decrease in both mean log wages and overall wage inequality being associated with education. However, the magnitude of these effects is reasonably small. For example, education effects can explain only about 4% of the observed reduction in the "10–50 gap". The effect of changes in the distribution of potential labor market experience on the quantiles of the wage distribution turns out to be roughly the opposite of that of education, and is thus rather small in magnitude as well.

Since union coverage is measured by a binary indicator, the corresponding FPPE is not point identified. The estimated identified sets turn out to be wide, thus making a precise quantification of the role of the decline in unionization difficult. For example, Table 3 shows that deunionization could have contributed anything between 0.004 to

Table 3: Decomposition Results: Inequality Measures

	Variance	Q90-Q10	Q90-Q50	Q50-Q10	Gini
Total Change	0.065	0.133	0.183	-0.050	0.020
Full Policy Effect	0.024	0.087	0.039	0.048	0.068
Partial Policy Effect:					
Education	-0.005	-0.010	-0.008	-0.002	-0.017
Experience	0.008	0.030	0.019	0.011	0.029
Union Coverage					
Upper bound	0.021	0.065	0.090	0.015	-0.001
Lower bound	-0.004	0.004	0.019	-0.054	-0.031

Note: Sampling variation of estimates is negligible due to large sample size.

0.065 to the observed change of 0.133 in the "90-10 gap". However, since the identified set does not include zero, it can be concluded that deunionization contributed to the rise in inequality, even though the precise magnitude remains unclear. Similarly, we see that deunionization increased inequality at the top-end of the wage distribution as measured by the "90-50 gap" (between 0.019 and 0.090 of the totally observed 0.183). Due to the width of the identified set, its role in the evolution of low-end wage inequality remains unclear. Our estimates suggest that deunionization alone could have shifted the "10–50" gap by anything between -0.054 and 0.015, thus allowing for both positive and negative influence.

As a final remark, we note that the bounds on the FPPE of unionization on the variance and the Gini coefficient given in Table 3 have been obtained by maximizing and minimizing the bounds given a fixed value for the mean in Theorem 3 over the estimated identified region of $\mu(F_Y^H)$. These estimated bounds are thus conservative. For the purpose of illustration, Figure 3 shows an estimate of the joint identification region of $\alpha_W(\text{Var})$ and $\alpha_W(\mu)$. The identified set has the shape of two intersecting parabolas. The distance between the upper and lower bound on the variance effect decreases when moving away from the center of the identified set of $\alpha_W(\mu)$, and vanishes at the boundary.

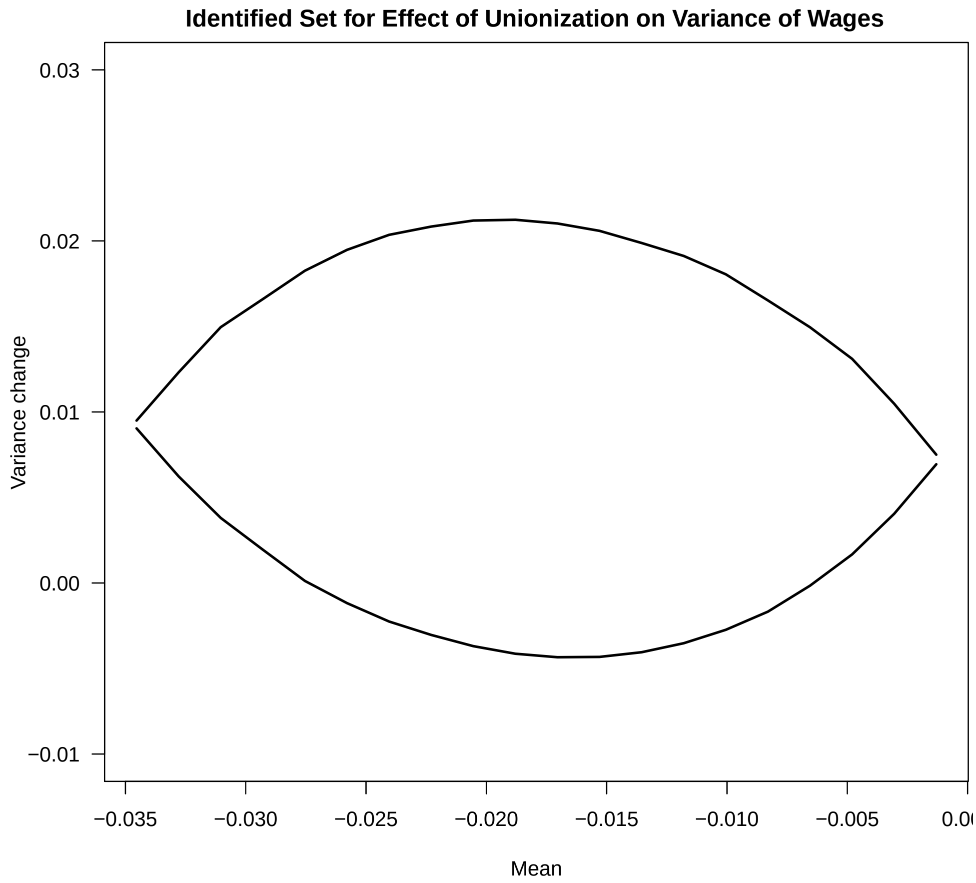

Figure 3: Joint identification region for $\alpha_W(\text{Var})$ and $\alpha_W(\mu)$ in case of union coverage.

References

Autor, D., L. Katz, and M. Kearney (2006): "The polarization of the US labor market," *American Economic Review*, 96, 189–194.

Barsky, R., J. Bound, K. Charles, and J. Lupton (2002): "Accounting for the black-white wealth gap," *Journal of the American Statistical Association*, 97, 663–673.

Blinder, A. (1973): "Wage discrimination: reduced form and structural estimates," *Journal of Human Resources*, 8, 436–455.

Carneiro, P., J. Heckman, and E. Vytlacil (2010): "Evaluating marginal policy changes and the average effect of treatment for individuals at the margin," *Econometrica*, 78, 377–394.

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

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

Chesher, A. (2003): "Identification in Nonseparable Models," *Econometrica*, 71, 1405–1441.

DiNardo, J., N. Fortin, and T. Lemieux (1996): "Labor market institutions and the distribution of wages, 1973-1992: A semiparametric approach," *Econometrica*, 64, 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, 609–633.

Firpo, S., N. Fortin, and T. Lemieux (2007): "Decomposing wage distributions using recentered influence function regressions," *Working Paper*.

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

——— (2010): "Decomposition Methods in Economics," *NBER Working Paper*.

Foresi, S. and F. Peracchi (1995): "The Conditional Distribution of Excess Returns: An Empirical Analysis." *Journal of the American Statistical Association*, 90, 451–466.

Goos, M. and A. Manning (2007): "Lousy and lovely jobs: The rising polarization of work in Britain," *The Review of Economics and Statistics*, 89, 118–133.

Gosling, A., S. Machin, and C. Meghir (2000): "The Changing Distribution of Male Wages in the UK," *Review of Economic Studies*, 67, 635–666.

Hoderlein, S. and E. Mammen (2007): "Identification of Marginal Effects in Nonseparable Models Without Monotonicity," *Econometrica*, 75, 1513–1518.

Imbens, G. and C. Manski (2004): "Confidence intervals for partially identified parameters," *Econometrica*, 72, 1845–1857.

Imbens, G. and W. Newey (2009): "Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity," *Econometrica*, 77, 1481–1512.

Lemieux, T. (2006): "Increasing residual wage inequality: Composition effects, noisy data, or rising demand for skill?" *American Economic Review*, 96, 461–498.

——— (2008): "The changing nature of wage inequality," *Journal of Population Economics*, 21, 21–48.

Machado, J. and J. Mata (2005): "Counterfactual decomposition of changes in wage distributions using quantile regression," *Journal of Applied Econometrics*, 20, 445–465.

Manski, C. (2003): *Partial identification of probability distributions*, Springer Verlag.

——— (2007): *Identification for prediction and decision*, Harvard University Press.

Manski, C. and E. Tamer (2002): "Inference on regressions with interval data on a regressor or outcome," *Econometrica*, 70, 519–546.

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

Melly, B. (2005): "Decomposition of differences in distribution using quantile regression," *Labour Economics*, 12, 577–590.

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

Newey, W. (1994): "The asymptotic variance of semiparametric estimators," *Econometrica*, 62, 1349–1382.

Oaxaca, R. (1973): "Male-female wage differentials in urban labor markets," *International Economic Review*, 14, 693–709.

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

Rüschendorf, L. (2009): "On the distributional transform, Sklar's theorem, and the empirical copula process," *Journal of Statistical Planning and Inference*, 139, 3921–3927.

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

Stock, J. (1989): "Nonparametric Policy Analysis," *Journal of the American Statistical Association*, 84, 567–575.

Stoye, J. (2009): "More on confidence intervals for partially identified parameters," *Econometrica*, 77, 1299–1315.

——— (2010): "Partial Identification of Spread Parameters," *Working Paper*.

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