DISCUSSION PAPER SERIES

Partial Distributional Policy Effects

Christoph Rothe

October 2011

Partial Distributional Policy Effects

Christoph Rothe
Toulouse School of Economics
and IZA

ABSTRACT

Partial Distributional Policy Effects[*]

In this paper, we propose a method to evaluate the effect of a counterfactual change in the unconditional 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.

Corresponding author:

Christoph Rothe
Toulouse School of Economics (GREMAQ)
21 Allée de Brienne
31000 Toulouse
France
E-mail: rothe@cict.fr

1. Introduction

In this paper, we introduce a new class of parameters called Partial Distributional Policy Effects (PPEs), which measure the partial (or *ceteris paribus*) effect of a counterfactual change in the unconditional distribution of a single component of a vector of explanatory variables X on some feature of the unconditional distribution of an outcome variable Y. The parameters are defined in the context of a general nonseparable model (e.g. Matzkin, 2003), and both fixed and marginal changes in the covariate distribution can be considered.[1] They provide direct answers to a wide range of interesting questions in applied economic analysis, such as how the quantiles of today's wage distribution would respond to a change in workers' age structure to that in, say, 1985 while holding all other characteristics of the labor force fixed, or how wage inequality would react to a marginal increase in the proportion of unionized workers, assuming again that all other features of the labor market remain constant.

The conceptional difficulties to define unconditional *ceteris paribus* effects are caused by the possible nonlinearity of the relationship between the distributions of Y and $X = (W, Z)$. If we were to consider a simple linear model, and restrict our attention to the mean of Y, our PPEs would reduce to simple functions of regression coefficients. When $Y = \beta_0 + \beta_1 W + \beta_2 Z + \varepsilon$, and we write $\mathbb{E}(W) = \mu_W$, a change in the distribution of W to another one with mean μ^* mechanically increases the expectation of the outcome variable by $(\mu^* - \mu_W)\beta_1$, whereas the effect of a marginal increase in the location of W is easily seen to be equal to $\partial \mathbb{E}(Y)/\partial \mu_W = \beta_1$.

Our PPEs generalize these simple ideas in several important direction. Specifically, they possess three important features, whose joint occurrence distinguishes them from other concepts proposed in the literature. First, they allow for general changes in the covariate distribution, and not only for 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 outcome distribution, such as its variance, higher moments,

---

[1]Throughout the paper, we refer to the marginal distribution of a single component of a random vector as its "unconditional" distribution. The term "marginal" is used in the sense of "infinitesimal" only.

quantiles, quantile differences, or Gini coefficient, and not only on the mean.

We formalize the ambiguous notion of a *ceteris paribus* change in the distribution of one covariate by introducing a rank invariance condition. That is, we construct a counterfactual experiment in such a way that the joint distribution of individuals' respective covariate ranks remains constant, thus preserving the dependence structure of individuals' observable characteristics. This approach is equivalent to holding the copula of the covariate distribution constant. We show that, under a conditional exogeneity condition, the unconditional distribution of the outcome variable after such a counterfactual experiment can be obtained by integrating the conditional cumulative distribution function (CDF) of Y given X with respect to the counterfactual covariate distribution, and one can thus directly calculate distributional features of interest. Comparing these features to their counterparts in the original outcome distribution yields our PPEs. We also discuss both parametric and nonparametric sample analogue estimators based on this result, which are straightforward to implement.

A complication arises for discrete covariates. In this case, the rank of an individual in the respective unconditional distribution is not uniquely determined by the data, but only upper and lower bounds can be obtained. 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.[2] This finding should not be seen as a weakness of our approach, but points to the difficulties to define unconditional *ceteris paribus* effects in general nonlinear models. Following the literature on partially identified parameters (e.g. Manski, 2003, 2007), we derive bounds on the PPEs in the case of discrete covariates.

We see the class of PPEs as an important complement to other measures that are commonly used in policy analysis. Their advantage is that they offer a clear distinction of the effect of a change in the unconditional distribution of a covariate from that of a change in the composition of subgroups defined by the covariates in the population, while other methods mix those effects. A simple example is useful to understand this

---

[2]The inability to uniquely determine an individual's rank in the presence of discrete data is also the source of partial identification of structural functions in triangular systems with discrete endogenous covariates studied in Chesher (2005) and Jun, Pinkse, and Xu (2011).

point. Suppose we study a population of workers, our covariates consist of age and an indicator for union coverage, and our outcome variable is the hourly wage. If the wage setting function contains interaction effects, unionization can affect the unconditional distribution of wages in this setting essentially through two channels: i) the mere proportion of unionized workers, and ii) the age composition of unionized and non-unionized workers. Our definition of a PPE of a change in "union coverage" is via a counterfactual experiment that affects wages through the first channel only. A different approach, considered e.g. in DiNardo, Fortin, and Lemieux (1996), Altonji, Bharadwaj, and Lange (2008) or Chernozhukov, Fernandez-Val, and Melly (2009a) would be to change the conditional distribution of unionization given age. This would exert an effect through both channels, since it both affects the proportion and the age composition of unionized and non-unionized workers. One could also consider to change the proportion of unionized workers while holding the conditional distribution of age given unionization constant, as e.g. in Machado and Mata (2005). In such a thought experiment, the wage distribution would again be affected through both channels, and additionally through the change in the distribution of workers' age. Which of these three different approaches is the most appropriate one thus depends on the exact research question one is trying to answer. In our simple example, PPEs should be of interest whenever it is empirically relevant to distinguish the effect of *more* workers being covered by unions from the effect of *different* workers being covered by unions, holding all other features of the population constant.

This paper contributes to an extensive literature on the analysis of counterfactual distributions, surveyed in Firpo et al. (2010). The impact of fixed changes in the *entire* covariate distribution is studied for example by Stock (1989), DiNardo et al. (1996), Gosling, Machin, and Meghir (2000), Donald, Green, and Paarsch (2000), Barsky, Bound, Charles, and Lupton (2002), Machado and Mata (2005), Melly (2005), Chernozhukov et al. (2009a) and Rothe (2010). As pointed out by Firpo et al. (2010), this literature is closely related to the problem of estimating average treatment effects and missing data models under unconfoundedness, see e.g. Hahn (1998), Hirano, Imbens, and Ridder (2003), Firpo (2007) or Chen, Hong, and Tarozzi (2008). As already discussed above, DiNardo et al. (1996) and Altonji et al. (2008) study the effect of a change in the conditional distribution of a single covariate given the remaining ones, and Machado and Mata (2005)

consider the effect of a change in the unconditional distribution of one of the covariates while holding the conditional distribution of the remaining covariates constant. Firpo, Fortin, and Lemieux (2009) study 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. 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. Our PPEs could also be used to construct an extension of the popular Oaxaca-Blinder procedure (Oaxaca, 1973; Blinder, 1973) to decompose intra-group differences in means to nonlinear models and general features of the outcome distribution. Such an approach would have certain advantages over a procedure with a similar aim described by Firpo et al. (2010), which involves approximating the effect of a change in the unconditional distribution of a covariate by a linear function of the change in the mean of the covariate. This approximation might not be accurate for distributional features of the outcome variable other than the mean.

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. In Section 5, we present the results of a small-scale empirical application. Section 6 concludes. All proofs are collected in Appendix A. Further details are discussed in Appendix B–D.

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 and Q_Y denote the CDF and the quantile function of the unconditional distribution of Y, respectively, 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 $(d-1)$-dimensional vector of remaining covariates. We then write the observed covariates X in terms of their unconditional quantile functions and a vector $U = (U_1, \ldots, U_d)$ of standard uniformly distributed latent variables, i.e.

$$X = (Q_W(U_1), Q_{Z_1}(U_2), \ldots, Q_{Z_{d-1}}(U_d)) \tag{2.2}$$

for some $U_i \sim U[0,1]$ and $i = 1, \ldots, d$. 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 $X_H = (H^{-1}(U_1), Q_{Z_1}(U_2), \ldots, Q_{Z_{d-1}}(U_d)) = (H^{-1}(U_1), Z)$ 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 approaches F_W from a particular direction (chosen by the analyst) as $t \to 0$. Our definition can equivalently be understood as 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. Note that in contrast to e.g. Chernozhukov and Hansen (2005) or Torgovitsky (2011), we do not use the notion of rank invariance in the sense of a restriction on individuals' behavior to obtain identification, but to define the parameters of interest in the first place.

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

$$\alpha(\nu, W, H) = \nu(F_Y^H) - \nu(F_Y).$$

(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*

$$\beta(\nu, W, H_t) = \lim_{t \to 0} \frac{\nu(F_Y^t) - \nu(F_Y)}{t} = \left. \frac{\partial \nu(F_Y^t)}{\partial t} \right|_{t=0}$$

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.

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). We first obtain two representations of the counterfactual outcome distribution F_Y^H, which are useful for the identification analysis of fixed and marginal PPEs, respectively.

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) $F_Y^H(y) = \int F_{Y|X}(y|w, z)dC(H(w), F_{Z_1}(z_1), \dots, F_{Z_{d-1}}(z_{d-1}))$.*

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 $\mathcal{F}_Y^H$. This set could be defined equivalently via the copula representation in Lemma 1(ii).

We remark that our setup implies that there are no issues for learning the conditional CDF $F_{Y|X}$. This rules out settings where the realizations of the outcome variable are missing for some individuals, e.g. because we do not observe a wage when someone is unemployed. In such a case, our approach could be combined with nonseparable sample selection models that either restore point identification of $F_{Y|X}$ under certain conditions, or establish upper and lower bounds on this CDF (e.g. Blundell, Gosling, Ichimura, and Meghir, 2007; Arellano and Bonhomme, 2011; Melly and Huber, 2011).

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 $\alpha(\nu, W, H)$ 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 $\{0,1\}$. 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.[3] 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 that 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 normalized 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,[4] and define

---

[3]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 $\tilde{F}_W(1, V) = (1 - p_1) + Vp_1$.

[4]A random variable Q is said to be a normalized version of a random variable R conditional on a random vector S 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.

$V_\nu^L = 1 - V_\nu^U$. 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, that W is binary, and let ν be a linear functional. Then the FPPE $\alpha(\nu, W, H)$ is partially identified: we have that $\alpha^L(\nu, W, H) \leq \alpha(\nu, W, H) \leq \alpha^U(\nu, W, H)$, where $\alpha^r(\nu, W, H) = \nu(F^r) - \nu(F_Y)$ 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 is thus point identified irrespective of whether W is continuously distributed or not. Hence there are for example no identification issues related to discrete covariates in the classical Oaxaca-Blinder procedure, which is based on the linear model $m(w, z, e) = \alpha + \beta w + \gamma' z + e$.

One can show that Theorem 2 continues to hold without modifications if W is not binary but an arbitrary discrete random variable, as long as the function H^{-1} takes at most two values on the interval $(F_W(w-), F_W(w)]$ for every $w \in \mathbb{R}$. Moreover, a larger number of support points generally implies more narrow bounds, as with more support points the restriction $F_W(W-) < U_1 \leq F_W(W)$ on the value of the rank variable becomes more informative. Formally, if F_W and H are supported on $\{s_1, \ldots, s_k\}$ such that $\sup_{w \in \mathbb{R}}(F_W(w) - F_W(w-)) \to 0$ as $k \to \infty$, then clearly $\tilde{F}_W(W, V_\nu^U) - \tilde{F}_W(W, V_\nu^L) \overset{a.s.}{\to} 0$, which in turn implies that $\alpha^U(\nu, W, H) - \alpha^L(\nu, W, H) \to 0$ by dominated convergence. Treating a discrete random variable with many support points as continuous should thus be a reasonable approximation in many applications.

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, that W is binary, and that ν is either a D_1-parameter, D_2-parameter, or quantile contrast. Then the FPPE $\alpha(\nu, W, H)$ is partially identified: we have that $\alpha^L(\nu, W, H) \leq \alpha(\nu, W, H) \leq \alpha^U(\nu, W, H)$, where the upper and lower bounds are given as follows:*

(i) *For ν a D_1-parameter, we have $\alpha^r(\nu, W, H) = \nu(G^r) - \nu(F_Y)$ 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 $G_{\bar{\mu}}^U$ and $G_{\bar{\mu}}^L$ be the unique compressed and dispersed distributions (relative to G^U and G^L) with expectation $\mu(G_{\bar{\mu}}^L) = \mu(G_{\bar{\mu}}^U) = \bar{\mu}$. Then we have $\alpha^r(\nu, W, H) = \nu(G_{\bar{\mu}}^r) - \nu(F_Y)$ for $r \in \{U, L\}$.*

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

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.[5] 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.

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[6] at F_Y with derivative ν', (c) the partial derivative $C_1 = \partial C / \partial u_1$ of the copula function C exists, and (d) the unconditional distribution of W is continuous. Then the MPPE $\beta(\nu, W, H_t)$ is identified:*

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

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

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

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

---

[5]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.

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

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 generally 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 again 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. $H_t(w) = \mathbb{I}\{0 \leq w < 1\}(F_W(0) - t) + \mathbb{I}\{w \geq 1\}$. Then it follows from direct calculations that

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

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 $R_{\mathcal{Z}} = \{(F_{Z_1}(z_1), \ldots, F_{Z_{d-1}}(z_{d-1})) : z \in \mathcal{Z}\}$, 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 (a)–(c) of Theorem 4 hold, and that W is binary. Then the MPPE $\beta(\nu, W, H_t)$ is partially identified: we have that $\beta^L(\nu, W, H_t) \leq \beta(\nu, W, H_t) \leq \beta^U(\nu, W, H_t)$, where $\beta^U(\nu, W, H_t) = \sup_{z \in \mathcal{Z}} g_\nu(z)$ and $\beta^L(\nu, W, H_t) = \inf_{z \in \mathcal{Z}} g_\nu(z)$.*

It is an immediate consequence of Theorem 5 that $\beta(\nu, W, H_t)$ 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 $\nu(F) = F^{-1}$ we have that $g_\nu(z) = -(F_{Y|W,Z}(Q_Y(\tau)|1, z) - F_{Y|W,Z}(Q_Y(\tau)|0, z))/f_Y(Q_Y(\tau))$, 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 $\alpha(\nu, W, H)$ by the sample analogue estimator

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

Here $\hat{F}_Y^H(y) = (1/n)\sum_{i=1}^n \hat{F}_{Y|X}(y, \hat{H}^{-1}(\hat{F}_W(W_i)), Z_i)$, 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 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 function, or by a fully nonparametric CDF estimator,

e.g. a kernel estimator as in Rothe (2010).[7]

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 $\hat{\mathcal{A}}_W(\nu) = [\hat{\alpha}^L(\nu, W, H), \hat{\alpha}^U(\nu, W, H)]$, where

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

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 $\hat{G}_{\bar{\mu}}^L$, $\hat{G}_{\bar{\mu}}^U$, $\hat{G}_{\bar{m}}^L$ and $\hat{G}_\gamma^U$, 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 esti-

---

[7]While kernel estimators may lack precision in settings with a large number of covariates due to the "curse of dimensionality", they are typically easier to compute than many flexible parametric methods. For example, the method based on inverting a linear quantile regression requires solving a large number of optimization problems, whereas a kernel-based estimator can be shown to require only the computation of certain sample means (see Rothe, 2010, for details).

mated bounds under partial identification. While we do not present any formal results concerning efficiency, we conjecture that those could be obtained along the lines of e.g. Chen et al. (2008). 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 infimum operators in order to obtain median unbiased estimates. To be specific, the estimate of the identified set is given by $\hat{\mathcal{B}}_W(\nu) = [\hat{\beta}^L(\nu, W, H_t), \hat{\beta}^U(\nu, W, H_t)]$, where

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

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 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.

17

5. Empirical Application

In this section, we use our methodology for a short empirical application that investigates the role of composition effects in the polarization of the US labor market observed between 1985 and 2005 (Autor, Katz, and Kearney, 2006; Lemieux, 2008). During this period, wage inequality has been rising substantially in the top end of the wage distribution, but has slightly decreased in the bottom end (compare the left panel of Figure 1 below). We use our FPPEs to quantify to what extent changes in the unconditional distribution of various observable characteristics of the labor force, such as e.g. the decline in union coverage from 27% to 15%, or the increase in average age and education, can account for these observed patterns.

Our data set, extracted from the 1983-1985 and 2003-2005 Outgoing Rotation Group (ORG) supplements of the Current Population Survey (CPS), is identical to the one used by Firpo et al. (2010) (see also Lemieux (2006) for details on its construction). 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 our (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, marital status, and part-time status.

Using the procedures described in Section 4.1, we estimate the (identified set of the) FPPE $\alpha(\nu, W, H)$ for various functionals ν, with the role of W and W^* being taken by education, experience and union coverage, respectively. Results for other covariates are not reported for brevity. We treat the former two quantities as continuously distributed, and derive bounds only for the effect of union coverage. The conditional CDF of Y given X is estimated 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 richly parameterized logistic regression.[8] For reference, we also estimate the full

---

[8]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 or interaction terms, and $\beta(y)$ is a finite dimensional parameter indexed by $y \in \mathbb{R}$, that can be estimated by maximum likelihood. In addition to the covariates mentioned above, we use quadratic terms in education and experience and a full set of interaction terms
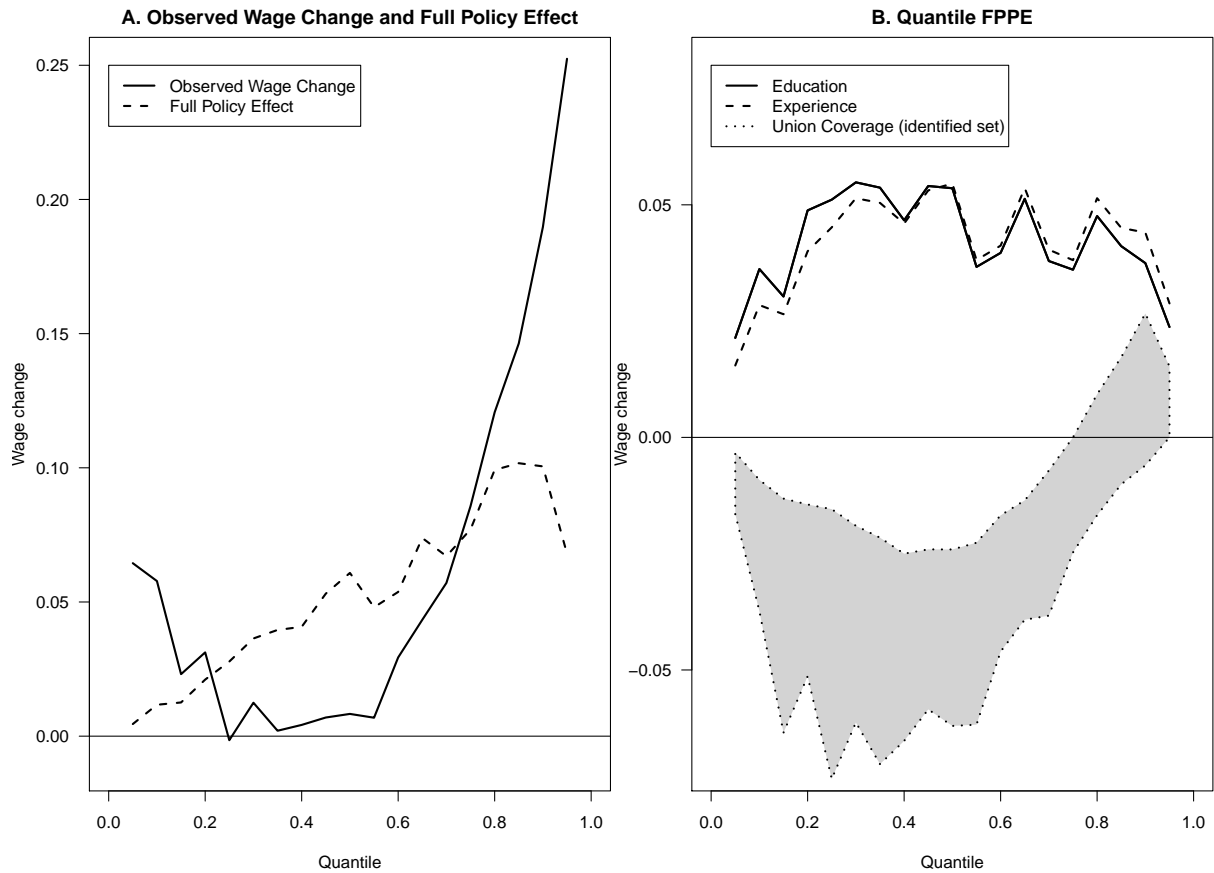
Figure 1: Relative change in τ-quantile of US male wages from 1985–2005 for $\tau \in (.05, .95)$. Left panel: Observed change (bold line) and estimated full distributional policy effect calculated using the method in Chernozhukov et al. (2009a) (dashed line). Right panel: FPPE of changes in education (bold line); FPPE of changes in experience (dashed line); identified set of FPPE of change in unionization (shaded area).

distributional policy effect of a change in the joint covariate distribution from that in 1983–1985 to that in 2003–2005 (e.g. Chernozhukov et al., 2009a; Rothe, 2010). Due to the large sample sizes, the sampling variation in our estimates is mostly negligible.

The left panel of Figure 1 compares estimates of the full distributional policy effect to the observed change 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 positive impact on each quantile, with the magnitude of the effect gradually increasing with the quantile under consideration. Accordingly, Table 1 shows that the full distributional policy effect

---

to estimate $F_{Y|X}$ (37 parameters in total).

Table 1: Estimated Fixed Partial Policy Effects of Changes in the Unconditional Distribution of Workers' Characteristics from that in 1983–1985 to that in 2003–2005 on the Distribution of Log Hourly Wages using CPS Data.

	Mean	Q10	Q50	Q90	Q90-Q10	Q90-Q50	Q50-Q10
Total Change	0.064	0.058	0.008	0.190	0.132	0.181	-0.050
	(1.68e-3)	(3.24e-3)	(3.05e-3)	(3.07e-3)	(4.75e-3)	(3.72e-3)	(4.13e-3)
Full Policy Effect	0.057	0.012	0.061	0.101	0.089	0.040	0.049
	(1.42e-3)	(1.88e-3)	(2.06e-3)	(1.69e-3)	(2.51e-3)	(2.15e-3)	(2.19e-3)
Partial Policy Effect:							
Education	0.046	0.036	0.054	0.037	0.001	-0.012	0.017
	(0.71e-3)	(1.51e-3)	(1.14e-3)	(1.13e-3)	(1.77e-3)	(1.26e-3)	(1.63e-3)
Experience	0.045	0.029	0.055	0.044	0.016	-0.011	0.026
	(0.74e-3)	(1.37e-3)	(1.13e-3)	(1.02e-3)	(1.59e-3)	(1.35e-3)	(1.52e-3)
Union Coverage							
Upper bound	-0.004	-0.009	-0.024	0.027	0.064	0.089	0.013
	(0.51e-3)	(1.19e-3)	(1.25e-3)	(0.78e-3)	(1.31e-3)	(1.21e-3)	(1.50e-3)
Lower bound	-0.038	-0.037	-0.062	-0.006	0.003	0.018	-0.053
	(0.51e-3)	(1.19e-3)	(1.25e-3)	(0.78e-3)	(1.31e-3)	(1.21e-3)	(1.50e-3)

Note: Partial Policy Effects are reported for selected covariates only. Other covariates included in the model are dummies for race, marital status, and part-time status. Bootstrapped standard errors in parentheses ($B = 200$ bootstrap replications; sampling variation of estimates is mostly negligible to due large sample size).

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-shape the observed changes exhibit, which must thus be driven by changes in the structural wage functions.

We now consider the FPPEs of individual covariates. From the right panel of Figure 1, we see that FPPEs of both education and experience on the quantiles of the (real log) wage distribution exhibit a similar "inverse U-shape", which implies a right shift of the (real log) wage distribution, an increase in lower-end inequality, and a decrease in top-end inequality. This is confirmed by the results in Table 1, which e.g. show a decrease in "90-50 gap" and an increase in the "50-10 gap" being associated with both education and experience.

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, which makes a precise quantification of the role of the decline in unionization difficult. For example, Table 1 shows that changes in unionization alone can account for anything between 0.003 to 0.064 to the observed change of 0.132 in the "90-10 gap". On the other hand, we see that deunionization increased inequality at the top-end of the wage distribution as measured by the "90-50 gap", contributing between 0.018 and 0.089 to the totally observed change of 0.181. 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 "50-10" gap by anything between -0.053 and 0.013, thus allowing for both positive and negative influence.

6. Concluding Remarks

In this paper, we propose a framework to define and to evaluate the effect of a counterfactual change in the unconditional 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.

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 $-\infty$ and $+\infty$, respectively. 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. We only prove statement (i), as the proof of statement (ii) is analogous. For notational simplicity, we assume that $d = 2$. First, it follows from the differentiability of the copula function and the functional form of H_t that

$$\partial_t\, C(H_t(w), F_Z(z))|_{t=0} = C_1(H(w), F_Z(z))\partial_t\, H_t(w)|_{t=0}$$
$$= C_1(H(w), F_Z(z))(G_W(u) - F_W(w)).$$

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

$$\lim_{t\to 0} \frac{F_Y^t(y) - F_Y(y)}{t} = \int F_{Y|X}(y|w, x) d\left(\partial_t\, C(H_t(w), F_Z(z))|_{t=0}\right)$$
$$= \int F_{Y|X}(y|w, x) d\left(C_1(H(w), F_Z(z))(G_W(u) - F_W(w))\right).$$

Finally, Hadamard differentiability of ν implies that

$$\lim_{t\to 0} \frac{\nu(F_Y^t) - \nu(F_Y)}{t} = \nu'\left(\int F_{Y|X}(y|w, x) d\left(C_1(H(w), F_Z(z))(G_W(u) - F_W(w))\right)\right)$$
$$= \int \nu'(F_{Y|X}(y|w, x)) d\left(C_1(H(w), F_Z(z))(G_W(u) - F_W(w))\right)$$

where the last equality follows from the linearity of ν'. $\qquad\square$

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 $R_{\mathcal{Z}} = \{(F_{Z_1}(z_1), \ldots, F_{Z_{d-1}}(z_{d-1})) : z \in \mathcal{Z}\}$, where $\mathcal{Z}$ denotes the support of X. For the case that the dimension of (W, Z) is equal to two, i.e. $d = 2$, this follows from Theorem 2.2.7 in Nelsen (2006). The extension of his result to the general multivariate case is immediate.

Next, let $\mathcal{T} = \{T : T(z) = s(F_{Z_1}(z_1), \ldots, F_{Z_{d-1}}(z_{d-1})), s \in \mathcal{S}\}$. 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. $\quad\square$

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 non-decreasing 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 "$\overset{d}{\to}$" 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* $\mathbf{F}_Y(y) = \sqrt{n}(\hat{F}_Y(y) - F_Y(y))$, $\mathbf{F}_X(w, z) = \sqrt{n}(\hat{F}_X(w, z) - F_X(w, z))$, $\mathbf{F}_W(w) = \sqrt{n}(\hat{F}_W(w) - F_W(w))$, *and* $\mathbf{Q}_W^*(\tau) = \sqrt{n}(\hat{H}^{-1}(\tau) - H^{-1}(\tau))$. *Then*

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

in the space $\ell^\infty(\mathcal{Y}) \times \ell^\infty(\mathcal{Y}) \times \ell^\infty(\mathcal{X}) \times \ell^\infty(\mathcal{W}) \times \ell^\infty([0, 1])$, *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–5th 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 1 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:*

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

27

where

$$\mathbb{F}_{Y,A}^H(y) = \int \mathbb{F}_{Y|X}(y, H^{-1}(F_W(w)), z)dF_X(w, z),$$

$$\mathbb{F}_{Y,B}^H(y) = \int F_{Y|X}(y, H^{-1}(F_W(w)), z)d\mathbb{F}_X(w, z),$$

$$\mathbb{F}_{Y,C}^H(y) = \int \partial_w F_{Y|X}(y, H^{-1}(F_W(u)), z)$$
$$\times \left(\partial_\tau H^{-1}(F_W(u))\mathbb{F}_W(u) + \sqrt{\lambda}\mathbb{Q}_W^*(F_W(u)) \right) dF_X(u, z),$$

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 and F_Y with derivative ν_H' and ν', respectively. Then*

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

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 $\mathbf{G}_{Y,s}^H = \int \mathbb{F}_{Y|X}(y, S_\nu^s(w, z), z)dF_X(w, z)$ for $s \in \{U, L\}$, we have that*

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

in the space $\ell^\infty(\mathcal{Y}) \times \ell^\infty(\mathcal{Y}) \times \ell^\infty(\mathcal{Y}) \times \ell^\infty(\mathcal{X})$, and the right hand side is given by a mean zero Gaussian process. (ii) The function class $\{(w, z) \mapsto F_{Y|X}(y, S_\nu^s(w, z), z), y \in \mathbb{R}, s \in \{U, L\}\}$ is F_X-Donsker. (iii) $\Pr(\hat{S}_\nu^s(W, Z) = S_\nu^s(W, Z)) \to 1$ 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* $\mathbf{N}^s = \sqrt{n}(\hat{\alpha}^s(\nu, W, H) - \alpha^s(\nu, W, H))$, $s \in \{U, L\}$, *jointly converge in distribution to a Normal distribution with mean zero:*

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

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

$$\mathbb{F}_A^s(y) = \int \mathbb{F}_{Y|X}(y, S_\nu^s(w, z), z) dF_X(w, z),$$

$$\mathbb{F}_B^s(y) = \int F_{Y|X}(y, S_\nu^s(w, z), z) d\mathbb{F}_X(w, z),$$

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* $\mathbf{G}_{Y,s}^H = \int \mathbb{F}_{Y|X}(y, S_y^s(w, z), z) dF_X(w, z)$ *for* $s \in \{U, L\}$, *we have that*

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

in the space $\ell^\infty(\mathcal{Y}) \times \ell^\infty(\mathcal{Y}) \times \ell^\infty(\mathcal{Y}) \times \ell^\infty(\mathcal{X})$, *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 $\nu : \mathcal{F} \to \mathbb{R}$ is Hadamard differentiable at F_Y^H and F_Y with derivative ν_H' and ν', respectively. (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,*

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

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 from Theorem 8 and the Functional Delta Method.

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

i) If ν is a D_1-parameter, then

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

ii) If ν is a quantile contrast, then

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

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 to conduct uniform inference

31

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.[9] For any

---

[9]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

32

$p \in (0, 1)$, we define

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

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

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

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

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

$$\hat{\mathcal{B}}_W(\nu, p) = [\hat{\beta}^L(\nu, W, H_t; p), \hat{\beta}^U(\nu, W, H_t; p)].$$

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 $\hat{\beta}_W^U(\nu; 1/2)$ is a consistent and asymptotically median unbiased estimate of the upper bound $\beta^U(\nu, W, H_t)$ of the identified set, in the sense that

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

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 $\Delta_n^+ =$

---

negative version of the bound generating function

$\Delta_n \mathbb{I}\{\Delta_n > 0\}$, where $\Delta_n = \hat{\beta}_W^U(\nu; 1/2) - \hat{\beta}_W^L(\nu; 1/2)$, and $\hat{p}_n = \Phi(\tau_n \Delta_n^+)c$, where $\Phi(\cdot)$ is the standard normal CDF and

$$\tau_n = \log(n)/\max[\hat{\beta}^U(\nu, W, H_t; 3/4) - \hat{\beta}^U(\nu, W, H_t; 1/4), \hat{\beta}^L(\nu, W, H_t; 3/4) - \hat{\beta}^L(\nu, W, H_t; 1/4)].$$

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 $\beta(\nu, W, H_t)$ 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 considering 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 be 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 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.

References

ALTONJI, J., P. BHARADWAJ, AND F. LANGE (2008): "Changes in the characteristics of american youth: Implications for adult outcomes," *NBER Working paper*.

ARELLANO, M. AND S. BONHOMME (2011): "Quantile Selection Models," *Working paper*.

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.

BLUNDELL, R., A. GOSLING, H. ICHIMURA, AND C. MEGHIR (2007): "Changes in the distribution of male and female wages accounting for employment composition using bounds," *Econometrica*, 75, 323–363.

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.

CHEN, X., H. HONG, AND A. TAROZZI (2008): "Semiparametric efficiency in GMM models with auxiliary data," *The Annals of Statistics*, 36, 808–843.

CHERNOZHUKOV, V., I. FERNANDEZ-VAL, AND B. MELLY (2009a): "Inference on Counterfactual Distributions," *Working Paper*.

CHERNOZHUKOV, V. AND C. HANSEN (2005): "An IV model of quantile treatment effects," *Econometrica*, 73, 245–261.

CHERNOZHUKOV, V., S. LEE, AND A. ROSEN (2009b): "Intersection Bounds: Estimation and Inference," *Working Paper*.

CHESHER, A. (2005): "Nonparametric identification under discrete variation," *Econometrica*, 73, 1525–1550.

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. (2007): "Efficient semiparametric estimation of quantile treatment effects," *Econometrica*, 75, 259–276.

FIRPO, S., N. FORTIN, AND T. LEMIEUX (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.

GOSLING, A., S. MACHIN, AND C. MEGHIR (2000): "The Changing Distribution of Male Wages in the UK," *Review of Economic Studies*, 67, 635–666.

HAHN, J. (1998): "On the role of the propensity score in efficient semiparametric estimation of average treatment effects," *Econometrica*, 66, 315–331.

HIRANO, K., G. IMBENS, AND G. RIDDER (2003): "Efficient estimation of average treatment effects using the estimated propensity score," *Econometrica*, 71, 1161–1189.

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.

Jun, S., J. Pinkse, and H. Xu (2011): "Tighter bounds in triangular systems," *Journal of Econometrics*, 161, 122–128.

Lemieux, T. (2006): "Increasing residual wage inequality: Composition effects, noisy data, or rising demand for skill?" *The 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.

Melly, B. and M. Huber (2011): "Nonseparable Sample Selection Models," *Working paper*.

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*.

TORGOVITSKY, A. (2011): "Identification of Nonseparable Models with General Instruments," *Working paper*.

VAN DER VAART, A. (2000): *Asymptotic Statistics*, Cambridge University Press.