INFERENCE ON COUNTERFACTUAL DISTRIBUTIONS

VICTOR CHERNOZHUKOV[†]   IVÁN FERNÁNDEZ-VAL[§]   BLAISE MELLY[‡]

ABSTRACT. We develop inference procedures for counterfactual analysis based on regression methods. We analyze the effect of either changing the distribution of covariates, or changing the conditional distribution of the outcome given covariates, on the entire marginal distribution of the outcome. For both of these scenarios we derive functional central limit theorems for regression-based estimators of the status quo and counterfactual marginal distributions. This result allows us to construct simultaneous confidence sets for function-valued counterfactual effects, including the effects on the entire marginal distribution function, quantile function, and related functionals. These confidence sets can be used to test functional hypotheses such as no-effect, positive effect, or stochastic dominance. Our theory applies to general counterfactual changes and covers the main regression methods including classical, quantile, duration, and distribution regressions. We illustrate the results with an empirical application to wage decompositions using data for the United States. Results on distribution regression as a tool for modeling the entire conditional distribution, encompassing duration/transformation regression, and representing an alternative to quantile regression, are also of independent interest.

Key Words: Counterfactual distribution, decomposition analysis, policy analysis, quantile regression, distribution regression, duration/transformation regression, Hadamard differentiability of the counterfactual operator

1. INTRODUCTION

Counterfactual distributions are important ingredients in both decomposition analysis (e.g., Juhn, Murphy, and Pierce, 1993, DiNardo, Fortin, and Lemieux, 1996, Fortin, Lemieux, and Firpo, 2011) and policy analysis in economics (Stock, 1989, Heckman and Vytlacil, 2007). For example, we might be interested in predicting the effect of cleaning up a local hazardous waste site on the marginal distribution of housing prices (Stock, 1991). Or, we might be interested in decomposing differences in wage distributions between men and women into a discrimination effect, arising due to pay differences between men and women with the same characteristics, and a composition effect, arising due to differences in characteristics between men and women (Oaxaca, 1973, and Blinder, 1973). In either example, the key decomposition or policy effects are differences between observed and counterfactual distributions. Using econometric terminology, we can often think of a counterfactual distribution as the result of either a change in the distribution of a set of covariates X that determine the outcome variable of interest Y, or as a change in the relationship of the covariates with the outcome, i.e. a change in the conditional distribution of Y given X. Counterfactual analysis consists of evaluating the effects of such changes.

The main objective and contribution of this paper is to provide estimation and inference procedures for the entire marginal counterfactual distribution of Y and its functionals based on regression methods. Starting from regression estimators of the conditional distribution of the outcome given covariates and nonparametric estimators of the covariate distribution, we obtain uniformly consistent and asymptotically Gaussian estimators for functionals of the status quo and counterfactual marginal distributions of the outcome. Examples of these functionals include distribution functions, quantile functions, quantile effects, distribution effects, Lorenz curves, and Gini coefficients. We then construct confidence sets that take into account the sampling variation coming from the estimation of the conditional and covariate distributions. These confidence sets are uniform in the sense that they cover the entire functional with pre-specified probability and can be used to test functional hypotheses such as no-effect, positive effect, or stochastic dominance.

Our analysis specifically targets and covers the regression methods for estimating conditional distributions most commonly used in empirical work, including classical, quantile, duration/transformation, and distribution regressions. We consider simple counterfactual scenarios consisting of marginal changes in the values of a given covariate, as well as more elaborate counterfactual scenarios consisting of general changes in the covariate distribution or in the conditional distribution of the outcome given covariates. For example, the changes in the covariate and conditional distributions can correspond to known transformations of these distributions in a population or to the distributions in different populations. This array of alternatives allows us to answer a wide variety of counterfactual questions such as the ones mentioned above.

This paper contains two sets of new theoretical results. First, we establish the validity of the estimation and inference procedures under two high-level conditions. The first condition requires

the first stage estimators of the conditional and covariate distributions to satisfy a functional central limit theorem. The second condition requires validity of the bootstrap for estimating the limit laws of the first stage estimators. Under the first condition, we derive functional central limit theorems for the estimators of the counterfactual functionals of interest, taking into account the sampling variation coming from the first stage. Under both conditions, we show that the bootstrap is valid for estimating the limit laws of the estimators of the counterfactual functionals. The key new theoretical result to all these results is the Hadamard differentiability of the counterfactual operator – that maps the conditional distributions and covariate distributions into the marginal counterfactual distributions – with respect to its arguments, which we establish in the paper (Lemma D.1). Given this key result, the other theoretical results above follow from the functional delta method. A convenient and important feature of these results is that they automatically imply estimation and inference validity of any existing or potential estimation method that obeys the two high-level conditions set forth above.

The second set of results deals with estimation and inference under primitive conditions in two leading regression methods. Specifically, we prove that the high-level conditions – functional central limit theorem and validity of bootstrap – hold for estimators of the conditional distribution based on quantile and distribution regression. In the process of proving these results we establish also some auxiliary results, which are of independent interest. In particular, we derive a functional central limit theorem and prove the validity of exchangeable bootstrap for the empirical coefficient process of distribution regression. We also prove the validity of the exchangeable bootstrap for the empirical coefficient process of quantile regression. Prior work by Hahn (1995) and Feng, He, and Hu (2011) showed bootstrap validity only for estimating pointwise laws of quantile regression coefficients. Note that the exchangeable bootstrap covers the empirical, weighted, subsampling, and m out of n bootstraps as special cases, which gives much flexibility to the practitioner.

This paper contributes to the previous literature on counterfactual analysis based on regression methods. Stock (1989) introduced integrated kernel regression-based estimators to evaluate the mean effect of policy interventions. Gosling, Machin, and Meghir (2000) and Machado and Mata (2005) proposed quantile regression-based estimators to evaluate distributional effects, but provided no econometric theory for these estimators. Our paper contributes to this literature in two ways. First, building on Foressi and Peracchi (1995), we develop the use of distribution regression as a tool for modeling and estimating the entire conditional distribution in counterfactual analysis. Distribution regression encompasses the Cox (1972) transformation/duration model as a special case, and represents a useful alternative to Koenker and Bassett (1978) quantile regression. Second, we provide limit distribution theory as well as inference tools for counterfactual estimators based on quantile and distribution regression. Moreover, our main results are generic and apply to any estimator of the conditional and covariate distributions that satisfy the conditions mentioned above, including classical regression (Juhn, Murphy and Pierce, 1993), flexible duration regression (Donald, Green and Paarsch, 2000), and other potential approaches.

An alternative approach to counterfactual analysis, which is not covered by our theoretical results, consists in re-weighting the observations using the propensity score, in the spirit of Horvitz and Thompson (1952). For instance, DiNardo, Fortin, and Lemieux (1996) apply this idea to estimate counterfactual densities, Firpo (2007) to quantile treatment effects, and Donald and Hsu (2012) to the distribution and quantile functions of potential outcomes. Under correct specification, the regression and the weighting approaches are equally valid. In particular, if we use saturated specifications for the propensity score and conditional distribution, then both approaches lead to numerically identical results. An advantage of the regression approach is that the intermediate step—the estimation of the conditional model—is often of independent economic interest. For example, Buchinsky (1994) applies quantile regression to analyze the determinants of conditional wage distributions. This model nests the classical Mincer wage regression and is useful for decomposing changes in the wage distribution into factors associated with between-group and within-group inequality.

We illustrate our estimation and inference procedures with a decomposition analysis of the evolution of the U.S. wage distribution, motivated by the in influential article by DiNardo, Fortin, and Lemieux (1996). We complement their analysis by using a wider range of techniques, providing standard errors for the estimates of the main effects, and extending the analysis to the entire distribution using simultaneous confidence bands. We also compare quantile and distribution regression and discuss the different choices that must be made to implement our estimators. Our results reinforce the importance of the decline in the real minimum wage and the minor role of de-unionization in explaining the increase in wage inequality during the 80s.

We organize the rest of the paper as follows. Section 2 presents our setting, the counterfactual distributions and effects of interest, and gives conditions under which these effects have a causal interpretation. In Section 3 we describe regression models for the conditional distribution, define our proposed estimation and inference procedures, and outline the main estimation and inference results. Section 4 contains the main theoretical results under simple high-level conditions, which cover a broad array of estimation methods. In Section 5 we verify the previous high-level conditions for the main estimators of the conditional distribution function—quantile and distribution regressions—under suitable primitive conditions. In Section 6 we present the empirical application, and in Section 7 we conclude with a summary of the main results and pointing out some possible directions of future research. In the Appendix, we include all the proofs and additional technical results. We give a consistency result for bootstrap confidence bands, a numerical example comparing quantile and distribution regression, and additional empirical results in the online supplemental material (Chernozhukov, Fernandez-Val, and Melly, 2012).

2. The Setting for Counterfactual Analysis

2.1. **Counterfactual distributions.** In order to motivate the analysis, let us first set up a simple running example. Suppose we would like to analyze the wage differences between men and

women. Let 0 denote the population of men and 1 the population of women. Y_j denotes wages and X_j denotes job market-relevant characteristics affecting wages for populations $j = 0$ and $j = 1$. The conditional distribution functions $F_{Y_0|X_0}(y|x)$ and $F_{Y_1|X_1}(y|x)$ describe the stochastic assignment of wages to workers with characteristics x, for men and women, respectively. Let $F_{Y\langle 0|0\rangle}$ and $F_{Y\langle 1|1\rangle}$ represent the observed distribution function of wages for men and women, and $F_{Y\langle 0|1\rangle}$ represent the counterfactual distribution function of wages that would have prevailed for women had they faced the men's wage schedule $F_{Y_0|X_0}$:

$$F_{Y\langle 0|1\rangle}(y) := \int_{\mathcal{X}_1} F_{Y_0|X_0}(y|x)dF_{X_1}(x).$$

The latter distribution is called counterfactual, since it does not arise as a distribution from any observable population. Rather, this distribution is constructed by integrating the conditional distribution of wages for men with respect to the distribution of characteristics for women. This quantity is well defined if $\mathcal{X}_0$, the support of men's characteristics, includes $\mathcal{X}_1$, the support of women's characteristics, namely $\mathcal{X}_1 \subseteq \mathcal{X}_0$.

The difference in the observed wage distributions between men and women can be decomposed in the spirit of Oaxaca (1973) and Blinder (1973) as follows:

$$F_{Y\langle 1|1\rangle} - F_{Y\langle 0|0\rangle} = [F_{Y\langle 1|1\rangle} - F_{Y\langle 0|1\rangle}] + [F_{Y\langle 0|1\rangle} - F_{Y\langle 0|0\rangle}],$$

where the first term in brackets is due to differences in the wage structure and the second term is a composition effect due to differences in characteristics. We can decompose similarly any functional of the observed wage distributions such as the quantile function or Lorenz curve into wage structure and composition effects. These counterfactual effects are well defined statistical parameters and are widely used in empirical analysis, e.g. the first term of the decomposition is a measure of gender discrimination. It is important to note that these effects do not necessarily have a causal interpretation without additional conditions. Section 2.3 provides sufficient conditions for such an interpretation to be valid. Thus, our theory covers both the descriptive decomposition analysis and the causal policy analysis, because the statistical objects – the counterfactual distributions and their functionals – are the same in either case.

In what follows we formalize these definitions and treat the more general case with several populations. We suppose that the populations are labeled by $k \in \mathcal{K}$, and that for each population k there is a random d_x-vector X_k of covariates and a random outcome variable Y_k. The covariate vector is observable in all populations, but the outcome is only observable in populations $j \in \mathcal{J} \subseteq \mathcal{K}$. Given observability, we can identify the covariate distribution F_{X_k} in each population $k \in \mathcal{K}$, and the conditional distribution $F_{Y_j|X_j}$ in each population $j \in \mathcal{J}$, as well as the corresponding conditional quantile function $Q_{Y_j|X_j}$.[1] Thus, we can associate each F_{X_k} with label k and each

---

[1] The inference theory of Section 4 does not rely on observability of X_k and (X_j, Y_j), but it only requires that F_{X_k} and $F_{Y_j|X_j}$ are identified and estimable at parametric rates. In principle, F_{X_k} and $F_{Y_j|X_j}$ can correspond to distributions of latent random variables. For example, $F_{Y_j|X_j}$ might be the conditional distribution of an outcome that we observe censored due to top coding, or it might be a structural conditional function identified

$F_{Y_j|X_j}$ with label j. We denote the support of X_k by $\mathcal{X}_k \subseteq \mathbb{R}^{d_x}$ and the region of interest for Y_j by $\mathcal{Y}_j \subseteq \mathbb{R}$. [2] We assume for simplicity that the number of populations, $|\mathcal{K}|$, is finite. Further, we define $\mathcal{Y}_j \mathcal{X}_j = \{(y,x) : y \in \mathcal{Y}_j, x \in \mathcal{X}_j\}$, $\mathcal{Y} \mathcal{X} \mathcal{J} = \{(y,x,j) : (y,x) \in \mathcal{Y}_j \mathcal{X}_j, j \in \mathcal{J}\}$, and generate other index sets by taking Cartesian products, e.g., $\mathcal{J} \mathcal{K} = \{(j,k) : j \in \mathcal{J}, k \in \mathcal{K}\}$.

Our main interest lies in the counterfactual distribution and quantile functions created by combining the conditional distribution in population j with the covariate distribution in population k, namely:

$$F_{Y\langle j|k\rangle}(y) := \int_{\mathcal{X}_k} F_{Y_j|X_j}(y|x) dF_{X_k}(x), \quad y \in \mathcal{Y}_j, \tag{2.1}$$

$$Q_{Y\langle j|k\rangle}(\tau) := F_{Y\langle j|k\rangle}^{\leftarrow}(\tau), \quad \tau \in (0,1), \tag{2.2}$$

where $F_{Y\langle j|k\rangle}^{\leftarrow}$ is the left-inverse function of $F_{Y\langle j|k\rangle}$ defined in Appendix A. In the definition (2.1) we assume the support condition:

$$\mathcal{X}_k \subseteq \mathcal{X}_j, \quad \text{for all } (j,k) \in \mathcal{J}\mathcal{K}, \tag{2.3}$$

which ensures that the integral is well defined. This condition is analogous to the overlap condition in treatment effect models with unconfoundedness (Rosenbaum and Rubin, 1983). In the gender wage gap example, it means that every female worker can be matched with a male worker with the same characteristics. If this condition is not met initially, we need to explicitly trim the supports and define the parameters relative to the common support.[3]

The counterfactual distribution $F_{Y\langle j|k\rangle}$ is the distribution function of the counterfactual outcome $Y\langle j|k\rangle$ created by first sampling the covariate X_k from the distribution F_{X_k} and then sampling $Y\langle j|k\rangle$ from the conditional distribution $F_{Y_j|X_j}(\cdot|X_k)$. This mechanism has a strong representation in the form[4]

$$Y\langle j|k\rangle = Q_{Y_j|X_j}(U|X_k), \text{ where } U \sim U(0,1) \text{ independently of } X_k \sim F_{X_k}. \tag{2.4}$$

This representation is useful for connecting counterfactual analysis with various forms of regression methods that provide models for conditional quantiles. In particular, conditional quantile models imply conditional distribution models through the relation:

$$F_{Y_j|X_j}(y|x) \equiv \int_{(0,1)} 1\{Q_{Y_j|X_j}(u|x) \leq y\} du. \tag{2.5}$$

---

by IV methods in a model with endogeneity. We focus on the case of observable random variables, because it is convenient for the exposition and covers our leading examples in Section 5. We briefly discuss extensions to models with endogeneity in the conclusion.

[2] We shall typically exclude tail regions of Y_j in estimation, as in Koenker (2005, p. 148).

[3] Specifically, given initial supports $\mathcal{X}_j^o$ and $\mathcal{X}_k^o$ such that $\mathcal{X}_k^o \not\subseteq \mathcal{X}_j^o$, we can set $\mathcal{X}_k = \mathcal{X}_j = (\mathcal{X}_k^o \cap \mathcal{X}_j^o)$. Then the covariate distributions are redefined over this support. See, e.g. Heckman, Ichimura, Smith, and Todd (1998), and Crump, Hotz, Imbens, and Mitnik (2009) for relevant discussions.

[4] This representation for counterfactuals was suggested by Roger Koenker in the context of quantile regression, as noted in Machado and Mata (2005).

In what follows, we define a *counterfactual effect* as the result of a shift from one counterfactual distribution $F_{Y\langle l|m\rangle}$ to another $F_{Y\langle j|k\rangle}$. Let $t = (j,k,l,m)$, for some $j, l \in \mathcal{J}$ and $k, m \in \mathcal{K}$. Then, we are interested in estimating and performing inference on the distribution and quantile effects

$$\Delta_t^{DE}(y) = F_{Y\langle j|k\rangle}(y) - F_{Y\langle l|m\rangle}(y) \text{ and } \Delta_t^{QE}(\tau) = Q_{Y\langle j|k\rangle}(\tau) - Q_{Y\langle l|m\rangle}(\tau),$$

as well as other functionals of the counterfactual distributions. For example, Lorenz curves, commonly used to measure inequality, are ratios of partial means to overall means

$$L(y, F_{Y\langle j|k\rangle}) = \int_{\mathcal{Y}_j} 1(t \leq y) t dF_{Y\langle j|k\rangle}(t) / \int_{\mathcal{Y}_j} t dF_{Y\langle j|k\rangle}(t),$$

defined for non-negative outcomes only, i.e. $\mathcal{Y}_j \subseteq [0, \infty)$. In general, the counterfactual effects take the form

$$\Delta_t(w) := \phi\left(F_{Y\langle j|k\rangle} : (j,k) \in \mathcal{JK}\right)(w). \tag{2.6}$$

This includes, as special cases, the previous distribution and quantile effects; Lorenz effects, with $\Delta_t(y) = L(y, F_{Y\langle j|k\rangle}) - L(y, F_{Y\langle l|m\rangle})$; Gini coefficients, with $\Delta_t = 1 - 2 \int_{\mathcal{Y}_j} L(F_{Y\langle j|k\rangle}, y) dy =: G_{Y\langle j|k\rangle}$; and Gini effects, with $\Delta_t = G_{Y\langle j|k\rangle} - G_{Y\langle l|m\rangle}$.

2.2. Types of counterfactuals effects.

Focusing on quantile effects as the leading functional of interest, we can isolate the following special cases of counterfactual effects (CE):

1) CE of changing the conditional distribution: $\qquad Q_{Y\langle j|k\rangle}(\tau) - Q_{Y\langle l|k\rangle}(\tau)$.
2) CE of changing the covariate distribution: $\qquad Q_{Y\langle j|k\rangle}(\tau) - Q_{Y\langle j|m\rangle}(\tau)$.
3) CE of changing the conditional and covariate distributions: $\quad Q_{Y\langle j|k\rangle}(\tau) - Q_{Y\langle l|m\rangle}(\tau)$.

In the gender wage gap example mentioned at the beginning of the section, the wage structure effect is a type 1 CE (with $j = 1$, $k = 1$, and $l = 0$), while the composition effect is an example of a type 2 CE (with $j = 0$, $k = 1$, and $m = 0$). In the wage decomposition application in Section 6 the populations correspond to time periods, the minimum wage is treated as a feature of the conditional distribution, and the covariates include union status and other worker characteristics. We consider type 1 CEs by sequentially changing the minimum wage and the wage structure. We also consider type 2 CEs by sequentially changing the components of the covariate distribution. The CE of simultaneously changing the conditional and covariate distributions are also covered by our theoretical results but are less common in applications.

While in the previous examples the populations correspond to different demographic groups or time periods, we can also create populations artificially by transforming status quo populations. This is especially useful when considering type 2 CE. Formally, we can think of X_k as being created through a known transformation of X_0 in population 0:

$$X_k = g_k(X_0), \quad \text{where } g_k : \mathcal{X}_0 \to \mathcal{X}_k. \tag{2.7}$$

This case covers, for example, adding one unit to the first covariate, $X_{1k} = X_{10} + 1$, holding the rest of the covariates constant. The resulting effect becomes the *unconditional* quantile regression, which measures the effect of a unit change in a given covariate component on the

unconditional quantiles of Y.[5] For example, this type of counterfactual is useful for estimating the effect of smoking on the marginal distribution of infant birth weights. Another example is a mean preserving redistribution of the first covariate implemented as $X_{1k} = (1-\alpha)E[X_{10}]+\alpha X_{10}$. These and more general types of transformation defined in (2.7) are useful for estimating the effect of a change in taxation on the marginal distribution of food expenditure, or the effect of cleaning up a local hazardous waste site on the marginal distribution of housing prices (Stock, 1991).

Even though the previous examples correspond to conceptually different thought experiments, our econometric analysis covers all of them.

2.3. **When counterfactual effects have a causal interpretation.** Under an assumption called conditional exogeneity, selection on observables or unconfoundedness (e.g., Rosenbaum and Rubin, 1983, Heckman and Robb, 1984, and Imbens, 2004), CE can be interpreted as causal effects. In order to explain this assumption and define causal effects, it is convenient to rely upon the potential outcome notation. Let $(Y_j^* : j \in \mathcal{J})$ denote a vector of potential outcome variables for various values of a policy, $j \in \mathcal{J}$, and let X be a vector of control variables or, simply, covariates.[6] Let J denote the random variable describing the realized policy, and $Y := Y_J^*$ the realized outcome variable. When the policy J is not randomly assigned, it is well known that the distribution of observed outcome Y conditional on $J = j$, i.e. the distribution of $Y \mid J = j$, may differ from the distribution of Y_j^*. However, if J is randomly assigned conditional on the control variables X—i.e. if the conditional exogeneity assumption holds—then the distributions of $Y \mid X, J = j$ and $Y_j^* \mid X$ agree. In this case the observable conditional distributions have a causal interpretation, and so do the counterfactual distributions generated from these conditionals by integrating out X.

To explain this point formally, let $F_{Y_j^* \mid J}(y \mid k)$ denote the distribution of the potential outcome Y_j^* in the population with $J = k \in \mathcal{J}$. The causal effect of exogenously changing the policy from l to j on the distribution of the potential outcome in the population with realized policy $J = k$, is

$$F_{Y_j^* \mid J}(y \mid k) - F_{Y_l^* \mid J}(y \mid k).$$

In the notation of the previous sections, the policy J corresponds to an indicator for the population labels $j \in \mathcal{J}$, and the observed outcome and covariates are generated as $Y_j = Y \mid J = j$,

---

[5]The resulting notion of unconditional quantile regression is related but strictly different from the notion introduced by Firpo, Fortin and Lemieux (2009). The latter notion measures a first order approximation to such an effect, whereas the notion described here measures the exact size of such an effect on the unconditional quantiles. When the change is relatively small, the two notions coincide approximately, but generally they can differ substantially.

[6]We use the term policy in a very broad sense, which could include any program or treatment. The definition of potential outcomes relies implicitly on a notion of manipulability of the policy via some thought experiment. Here there are different views about whether such thought experiment should be implementable or could be a purely mental act, see e.g., Rubin (1978) and Holland (1986) for the former view, and Heckman (1998, 2008) and Bollen and Pearl (2012) for the latter view. We exclude general equilibrium effects in the definition of potential outcomes.

and $X_k = X \mid J = k$.[7] The lemma given below shows that under conditional exogeneity, for any $j, k \in \mathcal{J}$ the counterfactual distribution $F_{Y\langle j|k\rangle}(y)$ exactly corresponds to $F_{Y_j^*|J}(y \mid k)$, and hence the causal effect of exogenously changing the policy from l to j in the population with $J = k$ corresponds to the CE of changing the conditional distribution from l to j, i.e.,

$$F_{Y_j^*|J}(y \mid k) - F_{Y_l^*|J}(y \mid k) = F_{Y\langle j|k\rangle}(y) - F_{Y\langle l|k\rangle}(y), \quad k \in \mathcal{J}.$$

Lemma 2.1 (Causal interpretation for counterfactual distributions)**.** *Suppose that*

$$(Y_j^* : j \in \mathcal{J}) \perp\!\!\!\perp J \mid X, \ a.s., \tag{2.8}$$

where $\perp\!\!\!\perp$ *denotes independence. Under (2.3) and (2.8),*

$$F_{Y\langle j|k\rangle}(\cdot) = F_{Y_j^*|J}(\cdot \mid k), \quad j, k \in \mathcal{J}.$$

The CE of changing the covariate distribution, $F_{Y\langle j|k\rangle}(y) - F_{Y\langle j|m\rangle}(y)$, also has a causal interpretation as the policy effect of changing exogenously the covariate distribution from F_{X_m} to F_{X_k} under additional conditions. Such a policy effect arises, for example, in Stock (1991)'s analysis of the impact of cleaning up a hazardous site on housing prices. Here, the distance to the nearest hazardous site is one of the characteristics, X, that affect the price of a house, Y, and the cleanup changes the distribution of X, say, from F_{X_m} to F_{X_k}. The assumption for causality is that the cleanup does not alter the hedonic pricing function $F_{Y_m|X_m}(y|x)$, which describes the stochastic assignment of prices y to houses with characteristics x. We do not discuss explicitly the potential outcome notation and the formal causal interpretation for this case.

3. MODELING CHOICES AND INFERENCE METHODS FOR COUNTERFACTUAL ANALYSIS

In this section we discuss modeling choices, introduce our proposed estimation and inference methods, and outline our results, without submersing into mathematical details. Counterfactual distributions in our framework have the form (2.1), so we need to model and estimate the conditional distributions $F_{Y_j|X_j}$ and covariate distributions F_{X_k}. As leading approaches for modeling and estimating $F_{Y_j|X_j}$ we shall use semi-parametric quantile and distribution regression methods. As a leading approach to estimating F_{X_k} we shall consider an unrestricted nonparametric method. Note that our proposal for using distribution regressions is new for counterfactual analysis, while our proposal for using quantile regressions builds on earlier work by Machado and Mata (2005).

3.1. **Regression models for conditional distributions.** The counterfactual distributions of interest depend on either the underlying conditional distribution, $F_{Y_j|X_j}$, or the conditional quantile function, $Q_{Y_j|X_j}$, through the relation (2.5). Thus, we can proceed by modeling and estimating either of these conditional functions. There are several principal approaches to carry out

---

[7]The notation $Y_j = Y \mid J = j$ designates that $Y_j = Y$ if $J = j$, and $X_k = X \mid J = k$ designates that $X_k = X$ if $J = k$.

these tasks, and our asymptotic inference theory covers these approaches as leading special cases. In this section we drop the dependence on the population index j to simplify the notation.

1. Conditional quantile models. Classical regression is one of the principal approaches to modeling and estimating conditional quantiles. The classical location-shift model takes the linear-in-parameters form: $Y = P(X)'\beta + V$, $V = Q_V(U)$, where $U \sim U(0,1)$ is independent of X, $P(X)$ is a vector of transformations of X such as polynomials or B-splines, and $P(X)'\beta$ is a location function such as the conditional mean. The additive disturbance V has unknown distribution and quantile functions F_V and Q_V. The conditional quantile function of Y given X is $Q_{Y|X}(u|x) = P(X)'\beta + Q_V(u)$, and the corresponding conditional distribution is $F_{Y|X}(y|x) = F_V(y - P(X)'\beta)$. This model, used in Juhn, Murphy and Pierce (1993), is parsimonious but restrictive, since no matter how flexible $P(X)$ is, the covariates impact the outcome only through the location. In applications this model as well as its location-scale generalizations are often rejected, so we cannot recommend its use without appropriate specification checks.

A major generalization and alternative to classical regression is quantile regression, which is a rather complete method for modeling and estimating conditional quantile functions (Koenker and Bassett, 1978, Koenker, 2005).[8] In this approach, we have the general non-separable representation: $Y = Q_{Y|X}(U|X) = P(X)'\beta(U)$, where $U \sim U(0,1)$ is independent of X(Koenker, 2005, p. 59). We can back out the conditional distribution from the conditional quantile function through the integral transform:

$$F_{Y|X}(y|x) = \int_{(0,1)} 1\{P(x)'\beta(u) \leq y\}du, \ y \in \mathcal{Y}.$$

The main advantage of quantile regression is that it permits covariates to impact the outcome by changing not only the location or scale of the distribution but also its entire shape. Moreover, quantile regression is flexible in that by considering $P(X)$ that is rich enough, one could approximate the true conditional quantile function arbitrarily well, when Y has a smooth conditional density (Koenker, 2005, p. 53).

2. Conditional distribution models. A common way to model conditional distributions is through the Cox (1972) transformation model: $F_{Y|X}(y|x) = 1 - \exp(-\exp(t(y) - P(x)'\beta))$, where $t(\cdot)$ is an unknown monotonic transformation. This conditional distribution corresponds to the following location-shift representation: $t(Y) = P(X)'\beta + V$, where V has an extreme value distribution and is independent of X. In this model, covariates impact an unknown monotone transformation of the outcome only through the location. The role of covariates is therefore limited in an important way. Note, however, that since $t(\cdot)$ is unknown this model is not a special case of quantile regression.

---

[8]Quantile regression is one of most important methods of regression analysis in economics. For applications, including to counterfactual analysis, see, e.g., Buchinsky (1994), Chamberlain (1994), Abadie (1997), Gosling, Machin, and Meghir (2000), Machado and Mata (2005), Angrist, Chernozhukov, and Fernández-Val (2006), and Autor, Katz, and Kearney (2006b).

Instead of restricting attention to the transformation model for the conditional distribution, we advocate modelling $F_{Y|X}(y|x)$ separately for all thresholds $y \in \mathcal{Y}$, developing further the ideas set forth in Foresi and Peracchi (1995) and Han and Hausman (1990).[9] Namely, we propose considering the *distribution regression* model

$$F_{Y|X}(y|x) = \Lambda(P(x)'\beta(y)), \quad y \in \mathcal{Y}, \tag{3.1}$$

where Λ is a known link function and $\beta(\cdot)$ is an unknown function-valued parameter. We note that this specification includes the Cox (1972) model as a strict special case, but allows for much more flexible effect of the covariates. Indeed, to see the inclusion, we set the link function to be the complementary log-log link, $\Lambda(v) = 1 - \exp(-\exp(v))$, take $P(x)$ to include a constant as the first component, and let $P(x)'\beta(y) = t(y) - P(x)'\beta$, so that only the first component of $\beta(y)$ varies with the threshold y. To see the greater flexibility of (3.1), we note that (3.1) allows all components of $\beta(y)$ to vary with y.

The fact that distribution regression with a complementary log-log link nests the Cox model leads us to consider this specification as an important reference point. Other useful link functions include the logit, probit, linear, log-log, and Gosset functions (see Koenker and Yoon, 2009, for the latter). We also note that the distribution regression model is flexible in the sense that, for any given link function Λ, we can approximate the conditional distribution function $F_{Y|X}(y|x)$ arbitrarily well by using a rich enough $P(X)$.[10] Thus, the choice of the link function is not important for sufficiently rich $P(X)$.

Comparison. It is important to compare and contrast the quantile regression and distribution regression models. Just like quantile regression generalizes location regression by allowing the slope coefficients $\beta(u)$ to depend on the quantile index u, distribution regression generalizes transformation (duration) regression by allowing the slope coefficients $\beta(y)$ to depend on the threshold index y. Both models therefore generalize important classical models and are semi-parametric because they have infinite-dimensional parameters $\beta(\cdot)$. When the specification of $P(X)$ is saturated, the quantile regression and distribution regression models coincide.[11] When the specification of $P(X)$ is not saturated, distribution and quantile regression models may differ substantially and are not nested. Accordingly, the model choice cannot be made on the basis of generality.

Both models are flexible in the sense that by allowing for a sufficiently rich $P(X)$, we can approximate the conditional distribution arbitrarily well. However, linear-in-parameters quantile

---

[9]Foresi and Peracchi (1995) propose estimating the conditional distribution by a logit model for several values of y. Previously, Han and Hausman (1990) considered an ordered logit specification. One of the main contributions of our paper is to extend this idea by developing distribution regression as a model for the entire conditional distribution function and deriving the corresponding limit theory for the distribution regression process.

[10]Indeed, let $P(X)$ denote the first p components of a basis in $L^2(\mathcal{X}, P)$. Suppose that $\Lambda^{-1}(F_{Y|X}(y|X)) \in L^2(\mathcal{X}, P)$ and $\lambda(t) = \partial\Lambda(t)/\partial t$ is bounded above by $\bar{\lambda}$. Then, for some $\beta(y)$ depending on p, $\delta_p = E\left[\Lambda^{-1}(F_{Y|X}(y|X)) - P(X)'\beta(y)\right]^2 \to 0$ as p grows, so that $E\left[F_{Y|X}(y|X) - \Lambda\left(P(X)'\beta(y)\right)\right]^2 \le \bar{\lambda}\delta_p \to 0$.

[11]For example, when $P(X)$ contains indicators of all points of support of X, if the support of X is finite.

regression is only flexible if Y has a smooth conditional density, and may provide a poor approximation to the conditional distribution otherwise, e.g. when Y is discrete or has mass points, as it happens in our empirical application. In sharp contrast, distribution regression does not require smoothness of the conditional density, since the approximation is done pointwise in the threshold y, and thus handles continuous, discrete, or mixed Y without any special adjustment. Another practical consideration is determined by the functional of interest. For example, we show in Remark 3.1 that the algorithm to compute estimates of the counterfactual distribution involves simpler steps for distribution regression than for quantile regression, whereas this computational advantage does not apply to the counterfactual quantile function. Thus, in practice, we recommend researchers to choose one method over the other on the basis of empirical performance, specification testing, ability to handle complicated data situations, or the functional of interest. In section 6 we explain how these factors influence our decision in a wage decomposition application.

3.2. **Estimation of counterfactual distributions and their functionals.** The estimator of each counterfactual distribution is obtained by the plug-in-rule, namely integrating an estimator of the conditional distribution $\widehat{F}_{Y_j|X_j}$ with respect to an estimator of the covariate distribution $\widehat{F}_{X_k}$,

$$\widehat{F}_{Y\langle j|k\rangle}(y) = \int_{\mathcal{X}_k} \widehat{F}_{Y_j|X_j}(y|x)d\widehat{F}_{X_k}(x), \; y \in \mathcal{Y}_j, \; (j,k) \in \mathcal{JK}. \tag{3.2}$$

For counterfactual quantiles and other functionals, we also obtain the estimators via the plug-in rule:

$$\widehat{Q}_{Y\langle j|k\rangle}(\tau) = \widehat{F}^{r\leftarrow}_{Y\langle j|k\rangle}(\tau) \text{ and } \widehat{\Delta}_t(w) = \phi(\widehat{F}_{Y\langle j|k\rangle} : (j,k) \in \mathcal{JK})(w), \tag{3.3}$$

where $\widehat{F}^r_{Y\langle j|k\rangle}$ denotes the rearrangement of $\widehat{F}_{Y\langle j|k\rangle}$ if $\widehat{F}_{Y\langle j|k\rangle}$ is not monotone (see Chernozhukov, Fernandez-Val, and Galichon, 2010).[12]

Assume that we have samples $\{(Y_{ki}, X_{ki}) : i = 1, ..., n_k\}$ composed of i.i.d. copies of (Y_k, X_k) for all populations $k \in \mathcal{K}$, where Y_{ji} is observable only for $j \in \mathcal{J} \subseteq \mathcal{K}$. We estimate the covariate distribution F_{X_k} using the empirical distribution function

$$\widehat{F}_{X_k}(x) = n_k^{-1} \sum_{i=1}^{n_k} 1\{X_{ki} \leq x\}, \quad k \in \mathcal{K}. \tag{3.4}$$

To estimate the conditional distribution $F_{Y_j|X_j}$, we develop methods based on the regression models described in Section 3.1. The estimator based on distribution regression (DR) takes the

---

[12]If a functional ϕ_0 requires proper distribution functions as inputs, we assume that the rearrangement is applied before applying ϕ_0. Hence formally, to keep notation simple, we interpret the final functional ϕ as the composition of the original functional ϕ_0 with the rearrangement.

form:

$$\widehat{F}_{Y_j|X_j}(y|x) = \Lambda(P(x)'\widehat{\beta}_j(y)), \quad (y,x) \in \mathcal{Y}_j\mathcal{X}_j, \quad j \in \mathcal{J}, \tag{3.5}$$

$$\widehat{\beta}_j(y) = \arg\max_{b\in\mathbb{R}^p} \sum_{i=1}^{n_j} \Big[1\{Y_{ji} \le y\}\ln[\Lambda(P(X_{ji})'b)] + 1\{Y_{ji} > y\}\ln[1 - \Lambda(P(X_{ji})'b)] \Big], \tag{3.6}$$

where $p = \dim P(X_j)$. The estimator based on quantile regression (QR) takes the form:

$$\widehat{F}_{Y_j|X_j}(y|x) = \varepsilon + \int_\varepsilon^{1-\varepsilon} 1\{P(x)'\widehat{\beta}_j(u) \le y\}du, \quad (y,x) \in \mathcal{Y}_j\mathcal{X}_j, \quad j \in \mathcal{J}, \tag{3.7}$$

$$\widehat{\beta}_j(u) = \arg\min_{b\in\mathbb{R}^p} \sum_{i=1}^{n_j} [u - 1\{Y_{ji} \le P(X_{ji})'b\}][Y_{ji} - P(X_{ji})'b], \tag{3.8}$$

for some small constant $\varepsilon > 0$. The trimming by ε avoids estimation of tail quantiles (Koenker, 2005, p. 148), and is valid under the conditions set forth in Theorem 4.1.[13]

We provide additional examples of estimators of the conditional distribution function in the working paper version (Chernozhukov, Fernandez-Val and Melly, 2009). Also our conditions in Section 4 allow for various additional estimators of the covariate distribution.

To sum-up, our estimates are computed using the following algorithm:

Algorithm 1 (Estimation of counterfactual distributions and their functionals). *(i) Obtain estimates $\widehat{F}_{X_k}$ of the covariate distributions F_{X_k} using (3.4). (ii) Obtain estimates $\widehat{F}_{Y_j|X_j}$ of the conditional distribution using (3.5)–(3.6) for DR or (3.7)–(3.8) for QR. (iii) Obtain estimates of the counterfactual distributions, quantiles and other functionals via (3.2) and (3.3).* □

Remark 3.1. In practice, the quantile regression coefficients can be estimated on a fine mesh $\varepsilon = u_1 < ... < u_S = 1-\varepsilon$, with meshwidth δ such that $\delta\sqrt{n_j} \to 0$. In this case the final counterfactual distribution estimator is computed as: $\widehat{F}_{Y\langle j|k\rangle}(y) = \varepsilon + n_k^{-1}\delta \sum_{i=1}^{n_k}\sum_{s=1}^S 1\{P(X_{ki})'\widehat{\beta}(u_s) \le y\}$. For distribution regression, the counterfactual distribution estimator takes the computationally convenient form $\widehat{F}_{Y\langle j|k\rangle}(y) = n_k^{-1} \sum_{i=1}^{n_k} \widehat{F}_{Y_j|X_j}(y|X_{ki})$ that does not involve inversion, trimming, nor fine mesh approximation. □

3.3. **Inference.** The estimators of the counterfactual effects follow functional limit theorems under conditions that we will make precise in the next section. For example, the estimators of the counterfactual distributions satisfy

$$\sqrt{n}(\widehat{F}_{Y\langle j|k\rangle} - F_{Y\langle j|k\rangle}) \rightsquigarrow \bar{Z}_{jk}, \quad \text{jointly in } (j,k) \in \mathcal{JK},$$

where n is a sample size index (say, n denotes the sample size of population 0) and $\bar{Z}_{jk}$ are zero-mean Gaussian processes. We characterize the limit processes for our leading examples in Section 5, so that we can perform inference using standard analytical methods. However, for

---

[13]In our empirical example, we use $\varepsilon = .01$. Tail trimming seems unavoidable in practice, unless we impose stringent tail restrictions on the conditional density or use explicit extrapolation to the tails as in Chernozhukov and Du (2008).

ease of inference, we recommend and prove the validity of a general resampling procedure called the *exchangeable bootstrap* (e.g., van der Vaart and Wellner, 1996). This procedure incorporates many popular forms of resampling as special cases, namely the empirical bootstrap, weighted bootstrap, m out of n bootstrap, and subsampling. It is quite useful for applications to have all of these schemes covered by our theory. For example, in small samples, we might want to use the weighted bootstrap to gain good accuracy and robustness to "small cells", whereas in large samples, where computational tractability can be an important consideration, we might prefer subsampling.

In the rest of this section we briefly describe the exchangeable bootstrap method and its implementation details, leaving a more technical discussion of the method to Sections 4 and 5. Let $(w_{k1}, ..., w_{kn_k})$, $k \in \mathcal{K}$, be vectors of nonnegative random variables that satisfy Condition EB in Section 5. For example, $(w_{k1}, ..., w_{kn_k})$ is a multinomial vector with dimension n_k and probabilities $(1/n_k, \ldots, 1/n_k)$ in the empirical bootstrap. The exchangeable bootstrap uses the components of $(w_{k1}, ..., w_{kn_k})$ as random sampling weights in the construction of the bootstrap version of the estimators. Thus, the bootstrap version of the estimator of the counterfactual distribution is

$$\widehat{F}^*_{Y\langle j|k\rangle}(y) = \int_{\mathcal{X}_k} \widehat{F}^*_{Y_j|X_j}(y|x) d\widehat{F}^*_{X_k}(x), \quad y \in \mathcal{Y}_j, \quad (j,k) \in \mathcal{JK}. \tag{3.9}$$

The component $\widehat{F}^*_{X_k}$ is a bootstrap version of covariate distribution estimator. For example, if using the estimator of F_{X_k} in (3.4), set

$$\widehat{F}^*_{X_k}(x) = (n^*_k)^{-1} \sum_{i=1}^{n_k} w_{ki} 1\{X_{ki} \leq x\}, \quad x \in \mathcal{X}_k, \quad k \in \mathcal{K}, \tag{3.10}$$

for $n^*_k = \sum_{i=1}^{n_k} w_{ki}$. The component $\widehat{F}^*_{Y_j|X_j}$ is a bootstrap version of the conditional distribution estimator. For example, if using DR, set $\widehat{F}^*_{Y_j|X_j}(y|x) = \Lambda(P(x)'\widehat{\beta}^*_j(y))$, $(y,x) \in \mathcal{Y}_j\mathcal{X}_j$, $j \in \mathcal{J}$, for

$$\widehat{\beta}^*_j(y) = \arg\max_{b \in \mathbb{R}^p} \sum_{i=1}^{n_j} w_{ji} \Big[1\{Y_{ji} \leq y\} \ln[\Lambda(P(X_{ji})'b)] + 1\{Y_{ji} > y\} \ln[1 - \Lambda(P(X_{ji})'b)] \Big].$$

If using QR, set $\widehat{F}^*_{Y_j|X_j}(y|x) = \varepsilon + \int_\varepsilon^{1-\varepsilon} 1\{P(x)'\widehat{\beta}^*_j(u) \leq y\} du$, $(y,x) \in \mathcal{Y}_j\mathcal{X}_j$, $j \in \mathcal{J}$, for

$$\widehat{\beta}^*_j(u) = \arg\min_{b \in \mathbb{R}^p} \sum_{i=1}^{n_j} w_{ji}[u - 1(Y_{ji} \leq P(X_{ji})'b)][Y_{ji} - P(X_{ji})'b].$$

Bootstrap versions of the estimators of the counterfactual quantiles and other functionals are obtained by monotonizing $\widehat{F}^*_{Y\langle j|k\rangle}$ using rearrangement if required and setting

$$\widehat{Q}^*_{Y\langle j|k\rangle}(\tau) = \widehat{F}^{*\leftarrow}_{Y\langle j|k\rangle}(\tau) \text{ and } \widehat{\Delta}^*_t(w) = \phi\left(\widehat{F}^*_{Y\langle j|k\rangle} : (j,k) \in \mathcal{JK}\right)(w). \tag{3.11}$$

The following algorithm describes how to obtain an exchangeable bootstrap draw of a counterfactual estimator.

Algorithm 2 (Exchangeable bootstrap for estimators of counterfactual functionals). *(i) Draw a realization of the vectors of weights $(w_{k1}, ..., w_{kn_k})$, $k \in \mathcal{K}$, that satisfy Condition EB in Section 5. (ii) Obtain a realization of the bootstrap version $\widehat{F}^*_{X_k}$ of the covariate distribution estimator $\widehat{F}_{X_k}$ using (3.10). (iii) Obtain a realization of the bootstrap version $\widehat{F}^*_{Y_j|X_j}$ of the conditional distribution estimator $\widehat{F}_{Y_j|X_j}$ using the same regression method as for the estimator. (iv) Obtain a realization of the bootstrap versions of the estimators of the counterfactual distribution, quantiles, and other functionals via (3.9) and (3.11).* □

The exchangeable bootstrap distributions are useful to perform asymptotically valid inference on the counterfactual effects of interest. We focus on uniform methods that cover standard pointwise methods for real-valued parameters as special cases, and also allow us to consider richer functional parameters and hypotheses. For example, an asymptotic simultaneous $(1 - \alpha)$-confidence band for the counterfactual distribution $F_{Y\langle j|k\rangle}(y)$ over the region $y \in \mathcal{Y}_j$ is defined by the end-point functions

$$\widehat{F}^{\pm}_{Y\langle j|k\rangle}(y) = \widehat{F}_{Y\langle j|k\rangle}(y) \pm \widehat{t}_{1-\alpha}\widehat{\Sigma}_{jk}(y)^{1/2}/\sqrt{n}, \tag{3.12}$$

such that

$$\lim_{n\to\infty} \mathbb{P}\left\{F_{Y\langle j|k\rangle}(y) \in [\widehat{F}^-_{Y\langle j|k\rangle}(y), \widehat{F}^+_{Y\langle j|k\rangle}(y)] \text{ for all } y \in \mathcal{Y}_j\right\} = 1 - \alpha. \tag{3.13}$$

Here, $\widehat{\Sigma}(y)$ is a uniformly consistent estimator of $\Sigma(y)$, the asymptotic variance function of $\sqrt{n}(\widehat{F}_{Y\langle j|k\rangle}(y) - F_{Y\langle j|k\rangle}(y))$. In order to achieve the coverage property (3.13), we set the critical value $\widehat{t}_{1-\alpha}$ as a consistent estimator of the $(1 - \alpha)$-quantile of the Kolmogorov-Smirnov maximal t-statistic:

$$t = \sup_{y\in\mathcal{Y}_j} \sqrt{n}\widehat{\Sigma}(y)^{-1/2}|\widehat{F}_{Y\langle j|k\rangle}(y) - F_{Y\langle j|k\rangle}(y)|.$$

The following algorithm describes how to obtain uniform bands using exchangeable bootstrap:

Algorithm 3 (Uniform inference for counterfactual analysis). *(i) Using Algorithm 2, draw $\{\widehat{Z}^*_{jk,b} : 1 \leq b \leq B\}$ as i.i.d. realizations of $\widehat{Z}^*_{jk}(y) = \sqrt{n}(\widehat{F}^*_{Y\langle j|k\rangle}(y) - \widehat{F}_{Y\langle j|k\rangle}(y))$, for $y \in \mathcal{Y}_j$, $(j,k) \in \mathcal{JK}$. (ii) Compute a bootstrap estimate of $\Sigma(y)^{1/2}$ such as the bootstrap interquartile range rescaled with the normal distribution: $\widehat{\Sigma}(y)^{1/2} = (q_{.75}(y) - q_{.25}(y))/1.349$ for $y \in \mathcal{Y}_j$, where $q_p(y)$ is the p-th quantile of $\{\widehat{Z}^*_{jk,b}(y) : 1 \leq b \leq B\}$. (3) Compute realizations of the maximal t-statistic $\widehat{t}_b = \sup_{y\in\mathcal{Y}_j} \widehat{\Sigma}(y)^{-1/2}|\widehat{Z}^*_{jk,b}(y)|$ for $1 \leq b \leq B$. (iii) Form a $(1 - \alpha)$-confidence band for $\{F_{Y\langle j|k\rangle}(y) : y \in \mathcal{Y}_j\}$ using (3.12) setting $\widehat{t}_{1-\alpha}$ to the $(1 - \alpha)$-sample quantile of $\{\widehat{t}_b : 1 \leq b \leq B\}$.* □

We can obtain similar uniform bands for the counterfactual quantile functions and other functionals replacing $\widehat{F}^*_{Y\langle j|k\rangle}$ by $\widehat{Q}^*_{Y\langle j|k\rangle}$ or $\widehat{\Delta}^*_t$ and adjusting the indexing sets accordingly. If the sample size is large, we can reduce the computational complexity of step (i) of the algorithm by resampling the first order approximation to the estimators of the conditional distribution, by using subsampling, or by simulating the limit process $\bar{Z}_{jk}$ using multiplier methods (Barrett and Donald, 2003).

Our confidence bands can be used to test functional hypotheses about counterfactual effects. For example, it is straightforward to test no-effect, positive effect or stochastic dominance hypotheses by verifying whether the entire null hypothesis falls within the confidence band of the relevant counterfactual functional, e.g., as in Barrett and Donald (2003) and Linton, Song, and Whang (2010).[14]

Remark 3.2 (On Validity of Confidence Bands)**.** Algorithm 3 uses the rescaled bootstrap interquartile range $\widehat{\Sigma}(y)$ as a robust estimator of $\Sigma(y)$. Other choices of quantile spreads are also possible adjusting the normal rescaling factor accordingly. If $\Sigma(y)$ is bounded away from zero on the region $y \in \mathcal{Y}_j$, uniform consistency of $\widehat{\Sigma}(y)$ over $y \in \mathcal{Y}_j$ and consistency of the confidence bands follow from the consistency of bootstrap for estimating the law of the limit Gaussian process $\bar{Z}_{jk}$, shown in Sections 4 and 5, and Lemma 1 in Chernozhukov and Fernandez-Val (2005); see Appendix A of the supplemental material for details. The bootstrap standard deviation is a natural alternative estimator for $\Sigma(y)$, but its consistency requires the uniform integrability of $\{(\widehat{Z}_{jk}^*(y))^2 : y \in \mathcal{Y}_j\}$, which in turn requires additional technical conditions that we do not impose (see Kato, 2011). □

4. Inference Theory for Counterfactual Analysis under General Conditions

This section contains the main theoretical results of the paper. We state the results under simple high-level conditions, which cover a broad array of estimation methods. We verify the high-level conditions for the principal approaches – quantile and distribution regressions – in the next section. Throughout this section, n denotes a sample size index and all limits are taken as $n \to \infty$. We refer to Appendix A for additional notation.

4.1. **Theory under general conditions.** We begin by gathering the key modeling conditions introduced in Section 2.

Condition S. *(a) The condition (2.3) on the support inclusion holds, so that the counterfactual distributions (2.1) are well defined. (b) The sample size n_k for the k-th population is nondecreasing in the index n and $n/n_k \to s_k \in [0, \infty)$, for all $k \in \mathcal{K}$, as $n \to \infty$.*

We impose high-level regularity conditions on the following empirical processes:

$$\widehat{Z}_j(y,x) := \sqrt{n_j}(\widehat{F}_{Y_j|X_j}(y|x) - F_{Y_j|X_j}(y|x)) \text{ and } \widehat{G}_k(f) := \sqrt{n_k} \int f d(\widehat{F}_{X_k} - F_{X_k}),$$

indexed by $(y,x,j,k,f) \in \mathcal{Y}\mathcal{X}\mathcal{J}\mathcal{K}\mathcal{F}$, where $\widehat{F}_{Y_j|X_j}$ is the estimator of the conditional distribution $F_{Y_j|X_j}$, $\widehat{F}_{X_k}$ is the estimator of the covariate distribution F_{X_k}, and $\mathcal{F}$ is a function class specified below. We require that these empirical processes converge to well-behaved Gaussian processes.

---

[14]For further references and other approaches, see McFadden (1989), Klecan, McFadden, and McFadden (1991), Anderson (1996), Davidson and Duclos (2000), Abadie (2002), Chernozhukov and Fernandez-Val (2005), Linton, Massoumi, and Whang (2005), Chernozhukov and Hansen (2006), or Maier (2011), among others.

In what follows, we consider $\mathcal{Y}_j\mathcal{X}_j$ as a subset of $\overline{\mathbb{R}}^{1+d_x}$ with topology induced by the standard metric ρ on $\overline{\mathbb{R}}^{1+d_x}$. We also let $\lambda_k(f, \tilde{f}) = [\int(f - \tilde{f})^2 dF_{X_k}]^{1/2}$ be a metric on $\mathcal{F}$.

Condition D. *Let $\mathcal{F}$ be a class of measurable functions that includes $\{F_{Y_j|X_j}(y|\cdot) : y \in \mathcal{Y}_j, j \in \mathcal{J}\}$ as well as indicators of all rectangles in $\overline{\mathbb{R}}^{d_x}$. (a) In the metric space $\ell^\infty(\mathcal{YXJKF})^2$,*

$$(\widehat{Z}_j(y,x), \widehat{G}_k(f)) \rightsquigarrow (Z_j(y,x), G_k(f)),$$

as stochastic processes indexed by $(y, x, j, k, f) \in \mathcal{YXJKF}$. The limit process is a zero-mean tight Gaussian process, where Z_j a.s. has uniformly continuous paths with respect to ρ, and G_k a.s. has uniformly continuous paths with respect to the metric λ_k on $\mathcal{F}$. (b) The map $y \mapsto F_{Y_j|X_j}(y|\cdot)$ is continuous with respect to the metric λ_k for all $(j, k) \in \mathcal{JK}$.

Condition D requires that a uniform central limit theorem hold for the estimators of the conditional and covariate distributions. We verify Condition D for semi-parametric estimators of the conditional distribution function, such as quantile and distribution regression, under i.i.d. sampling assumption. For the case of duration/transformation regression, this condition follows from the results of Andersen and Gill (1982) and Burr and Doss (1993). For the case of classical (location) regression, this condition follows from the results reported in the working paper version (Chernozhukov, Fernandez-Val and Melly, 2009). We expect Condition D to hold in many other applied settings. The requirement $\widehat{G}_k \rightsquigarrow G_k$ on the estimated measures is weak and is satisfied when $\widehat{F}_{X_k}$ is the empirical measure based on a random sample, as in the previous section. Finally, we note that Condition D does not even impose the i.i.d sampling conditions, only that a functional central limit theorem is satisfied. Thus, Condition D can be expected to hold more generally, which may be relevant for time series applications.

Remark 4.1 (Technical aspects). Condition D does not impose compactness assumptions on the regions $\mathcal{Y}_j$ or $\mathcal{X}_k$ per se, but we shall impose compactness when we provide primitive conditions. The requirement $\widehat{G}_k \rightsquigarrow G_k$ holds not only for empirical measures but also for various smooth empirical measures; in fact, in the latter case the indexing class of functions $\mathcal{F}$ can be much larger than Glivenko-Cantelli or Donsker; see Radulovic and Wegkamp (2003) and Gine and Nickl (2008). □

Theorem 4.1 (Uniform limit theory for counterfactual distributions and quantiles). *Suppose that Conditions S and D hold. (1) Then,*

$$\sqrt{n}\left(\widehat{F}_{Y\langle j|k\rangle}(y) - F_{Y\langle j|k\rangle}(y)\right) \rightsquigarrow \bar{Z}_{jk}(y) \tag{4.1}$$

as a stochastic process indexed by $(y, j, k) \in \mathcal{YJK}$ in the metric space $\ell^\infty(\mathcal{YJK})$, where $\bar{Z}_{jk}$ is a tight zero-mean Gaussian process with continuous paths on $\mathcal{Y}_j$ defined by

$$\bar{Z}_{jk}(y) := \sqrt{s_j} \int Z_j(y,x) dF_{X_k}(x) + \sqrt{s_k} G_k(F_{Y_j|X_j}(y|\cdot)). \tag{4.2}$$

16

(2) If in addition $F_{Y\langle j|k\rangle}$ admits a positive continuous density $f_{Y\langle j|k\rangle}$ on an interval $[a, b]$ containing an ϵ-enlargement of the set $\{Q_{Y\langle j|k\rangle}(\tau) : \tau \in \mathcal{T}\}$ in $\mathcal{Y}_j$, where $\mathcal{T} \subset (0, 1)$, then

$$\sqrt{n}\left(\widehat{Q}_{Y\langle j|k\rangle}(\tau) - Q_{Y\langle j|k\rangle}(\tau)\right) \rightsquigarrow -\bar{Z}_{jk}(Q_{Y\langle j|k\rangle}(\tau))/f_{Y\langle j|k\rangle}(Q_{Y\langle j|k\rangle}(\tau)) =: V_{jk}(\tau), \qquad (4.3)$$

as a stochastic process indexed by $(\tau, j, k) \in \mathcal{TJK}$ in the metric space $\ell^\infty(\mathcal{TJK})$, where V_{jk} is a tight zero mean Gaussian process with continuous paths on $\mathcal{T}$.

This is the first main and new result of the paper. It shows that if the estimators of the conditional and covariate distributions satisfy a functional central limit theorem, then the estimators of the counterfactual distributions and quantiles also obey a functional central limit theorem. This result forms the basis of all inference results on counterfactual estimators.

As an application of the result above, we derive functional central limit theorems for distribution and quantile effects. Let $t = (j, k, l, m)$, $\mathcal{Y} \subseteq \mathcal{Y}_j \cap \mathcal{Y}_l$, $\mathcal{T} \subset (0, 1)$, and

$$\Delta_t^{DE}(y) = F_{Y\langle j|k\rangle}(y) - F_{Y\langle l|m\rangle}(y), \quad \widehat{\Delta}_t^{DE}(y) = \widehat{F}_{Y\langle j|k\rangle}(y) - \widehat{F}_{Y\langle l|m\rangle}(y),$$
$$\Delta_t^{QE}(\tau) = Q_{Y\langle j|k\rangle}(\tau) - Q_{Y\langle l|m\rangle}(\tau), \quad \widehat{\Delta}_t^{QE}(\tau) = \widehat{Q}_{Y\langle j|k\rangle}(\tau) - \widehat{Q}_{Y\langle l|m\rangle}(\tau).$$

Corollary 4.1 (Limit theory for quantile and distribution effects). *Under the conditions of Theorem 4.1, part 1,*

$$\sqrt{n}\left(\widehat{\Delta}_t^{DE}(y) - \Delta_t^{DE}(y)\right) \rightsquigarrow \bar{Z}_{jk}(y) - \bar{Z}_{lm}(y) =: S_t(y), \qquad (4.4)$$

as a stochastic process indexed by $y \in \mathcal{Y}$ in the space $\ell^\infty(\mathcal{Y})$, where S_t is a tight zero-mean Gaussian process with continuous paths. Under conditions of Theorem 4.1, part 2,

$$\sqrt{n}\left(\widehat{\Delta}_t^{QE}(\tau) - \Delta_t^{QE}(\tau)\right) \rightsquigarrow V_{jk}(\tau) - V_{lm}(\tau) =: W_t(\tau), \qquad (4.5)$$

as a stochastic process indexed by $\tau \in \mathcal{T}$ in the space $\ell^\infty(\mathcal{T})$, where W_t is a tight zero-mean Gaussian process with continuous paths.

The following corollary is another application of the result above. It shows that Hadamard-differentiable functionals also satisfy a functional central limit theorem. Examples include Lorenz curves and Lorenz effects, as well as real-valued parameters, such as Gini coefficients and Gini effects. Regularity conditions for Hadamard-differentiability of Lorenz and Gini functionals are given in Bhattacharya (2007).

Corollary 4.2 (Limit theory for smooth functionals). *Consider the parameter θ as an element of a parameter space $\mathbb{D}_\theta \subset \mathbb{D} = \times_{(jk)\in\mathcal{JK}}\ell^\infty(\mathcal{Y}_j)$, with $\mathbb{D}_\theta$ containing the true value $\theta_0 = (F_{Y\langle j|k\rangle} : (j, k) \in \mathcal{JK})$. Consider the plug-in estimator $\widehat{\theta} = (\widehat{F}_{Y\langle j|k\rangle} : (j, k) \in \mathcal{JK})$. Suppose $\phi(\theta)$, a functional of interest mapping $\mathbb{D}_\theta$ to $\ell^\infty(\mathcal{W})$, is Hadamard differentiable in θ at θ_0 tangentially to $\times_{(jk)\in\mathcal{JK}}C(\mathcal{Y}_j)$ with derivative $(\phi'_{jk} : (j, k) \in \mathcal{JK})$. Let $\Delta_t = \phi(\theta_0)$ and $\widehat{\Delta}_t = \phi(\widehat{\theta})$. Then, under the conditions of Theorem 4.1, part 1,*

$$\sqrt{n}\left(\widehat{\Delta}_t(w) - \Delta_t(w)\right) \rightsquigarrow \sum_{(j,k)\in\mathcal{JK}} (\phi'_{jk}\bar{Z}_{jk})(w) =: T(w), \qquad (4.6)$$

as a stochastic processes indexed by $w \in \mathcal{W}$ in $\ell^\infty(\mathcal{W})$, where $w \mapsto T(w)$ is a tight zero-mean Gaussian process.

4.2. **Validity of resampling and other simulation methods for counterfactual analysis.** As we discussed in Section 3.3, Kolmogorov-Smirnov type procedures offer a convenient and computationally attractive approach for performing inference on function-valued parameters using functional central limit theorems. A complication in our case is that the limit processes in (4.2)–(4.6) are non-pivotal, as their covariance functions depend on unknown, though estimable, nuisance parameters.[15] We deal with this non-pivotality by using resampling and simulation methods. An attractive result shown as part of our theoretical analysis is that the counterfactual operator is Hadamard differentiable with respect to the underlying conditional and covariate distributions. As a result, if bootstrap or any other method consistently estimates the limit laws of the estimators of the conditional and covariate distributions, it also consistently estimates the limit laws of the estimators of the counterfactual distributions and their smooth functionals. This convenient result follows from the functional delta method for bootstrap of Hadamard differentiable functionals.

In order to state the results formally, we follow the notation and definitions in van der Vaart and Wellner (1996). Let D_n denote the data vector and M_n be the vector of random variables used to generate bootstrap draws or simulation draws given D_n (this may depend on the particular resampling or simulation method). Consider the random element $\mathbb{Z}_n^* = \mathbb{Z}_n(D_n, M_n)$ in a normed space $\mathbb{D}$. We say that the bootstrap law of $\mathbb{Z}_n^*$ consistently estimates the law of some tight random element $\mathbb{Z}$ and write $\mathbb{Z}_n^* \rightsquigarrow_\mathbb{P} \mathbb{Z}$ in $\mathbb{D}$ if

$$\sup_{h \in \mathrm{BL}_1(\mathbb{D})} |E_{M_n} h(\mathbb{Z}_n^*) - Eh(\mathbb{Z})| \to_\mathbb{P} 0, \tag{4.7}$$

where $\mathrm{BL}_1(\mathbb{D})$ denotes the space of functions with Lipschitz norm at most 1 and E_{M_n} denotes the conditional expectation with respect to M_n given the data D_n.

Next, consider the processes $\widehat{\vartheta}(t) = (\widehat{F}_{Y_j|X_j}(y|x), \int f d\widehat{F}_{X_k})$ and $\vartheta(t) = (F_{Y_j|X_j}(y|x), \int f dF_{X_k})$, indexed by $t = (y, x, j, k, f) \in T = \mathcal{YXJKF}$, as elements of $\mathbb{E}_\vartheta = \ell^\infty(T)^2$. Condition D(a) can be restated as $\sqrt{n}(\widehat{\vartheta}_n - \vartheta) \rightsquigarrow \mathbb{Z}_\vartheta$ in $\mathbb{E}_\vartheta$, where $\mathbb{Z}_\vartheta$ denotes the limit process in Condition D(a). Let $\widehat{\vartheta}_n^*$ be the bootstrap draw of $\widehat{\vartheta}_n$. Consider the functional of interest $\phi = \phi(\vartheta)$ in the normed space $\mathbb{E}_\phi$, which can be either the counterfactual distribution and quantile functions considered in Theorem 4.1, the distribution or quantile effects considered in Corollary 4.1, or any of the functionals considered in Corollary 4.2. Denote the plug-in estimator of ϕ as $\widehat{\phi} = \phi(\widehat{\vartheta})$ and the corresponding bootstrap draw as $\widehat{\phi}^* = \phi(\widehat{\vartheta}^*)$. Let $\mathbb{Z}_\phi$ denote the limit law of $\sqrt{n}(\widehat{\phi} - \phi)$, as described in Theorem 4.1, Corollary 4.1, and Corollary 4.2.

Theorem 4.2 (Validity of resampling and other simulation methods for counterfactual analysis)**.**
Assume that the conditions of Theorem 4.1 hold. If $\sqrt{n}(\widehat{\vartheta}_n^* - \widehat{\vartheta}) \rightsquigarrow_\mathbb{P} \mathbb{Z}_\vartheta$ in $\mathbb{E}_\vartheta$, then $\sqrt{n}(\widehat{\phi}^* - \widehat{\phi}) \rightsquigarrow_\mathbb{P}$

---

[15]Similar non-pivotality issues arise in a variety of goodness-of-fit problems studied by Durbin and others, and are referred to as the Durbin problem by Koenker and Xiao (2002).

$\mathbb{Z}_\phi$ in $\mathbb{E}_\phi$. *In words, if the exchangeable bootstrap or any other simulation method consistently estimates the law of the limit stochastic process in Condition D, then this method also consistently estimates the laws of the limit stochastic processes (4.2)–(4.6) for estimators of counterfactual distributions, quantiles, distribution effects, quantile effects, and other functionals.*

This is the second main and new result of the paper. Theorem 2 shows that any resampling method is valid for estimating the limit laws of the estimators of the counterfactual effects, provided this method is valid for estimating the limit laws of the (function-valued) estimators of the conditional and covariate distributions. We verify the latter condition for our principal estimators in Section 5, where we establish the validity of exchangeable bootstrap methods for estimating the laws of function-valued estimators of the conditional distribution based on quantile regression and distribution regression processes. As noted in Remark 3.2, this result also implies the validity of the Kolmogorov-Smirnov type confidence bands for counterfactual effects under non-degeneracy of the variance function of the limit processes for the estimators of these effects; see Appendix A of the supplemental material for details.

5. Inference Theory for Counterfactual Analysis under Primitive Conditions

We verify that the high-level conditions of the previous section hold for the principal estimators of the conditional distribution functions, and so the various conclusions on inference methods also apply to this case. We also present new results on limit distribution theory for distribution regression processes and exchangeable bootstrap validity for quantile and distribution regression processes, which may be of a substantial independent interest. Throughout this section, we re-label $P(X)$ to X to simplify the notation. This entails no loss of generality when $P(X)$ includes X as a subset.

5.1. **Preliminaries on sampling.** We assume there are samples $\{(Y_{ki}, X_{ki}) : i = 1, ..., n_k\}$ composed of i.i.d. copies of (Y_k, X_k) for all populations $k \in \mathcal{K}$. The samples are independent across $k \in \mathcal{K}_0 \subseteq \mathcal{K}$. We shall call the case with $\mathcal{K} = \mathcal{K}_0$ the *independent samples* case. We assume that Y_{ji} is observable only for $j \in \mathcal{J} \subseteq \mathcal{K}_0$. The independent samples case arises, for example, in the wage decomposition application of Section 6. In addition, we may have transformation samples indexed by $k \in \mathcal{K}_t$ created via transformation of some "originating" samples $l \in \mathcal{K}_0$. For example, in the unconditional quantile regression mentioned in Section 2, we create a transformation sample by shifting one of the covariates in the original sample up by a unit.

The transformation and originating samples are dependent, and we need to account for this dependency in the limit theory for counterfactual estimators. In order to do so formally, we specify the relation of each k-th transformation sample, with index $k \in \mathcal{K}_t$, to an originating sample, with index $l(k) \in \mathcal{K}_0$, as follows: $(Y_{ki}, X_{ki}) = g_{l(k),k}(Y_{l(k),i}, X_{l(k),i}), i = 1, ..., n_k$, where $g_{l(k),k}$ is a known measurable transformation map, and $l : \mathcal{K}_t \to \mathcal{K}_0$ is the indexing function that

gives the index $l(k)$ of the sample from which the transformation sample k is originated. We also let $\mathcal{K} = \mathcal{K}_t \cup \mathcal{K}_0$. The main requirement on the map $g_{l(k),k} : \mathbb{R}^{d+1} \mapsto \mathbb{R}$ is that it preserves the Dudley-Pollard's (sufficient) condition for universal Donskerness: namely, given a class $\mathcal{F}$ of suitably measurable and bounded functions mapping a measurable subset of $\mathbb{R}^{d+1}$ to $\mathbb{R}$ that obeys Pollard's entropy condition, the class $\mathcal{F} \circ g_{l(k),k}$ continues to contain bounded and suitably measurable functions, and obeys Pollard's entropy condition.[16] For example, this holds if $g_{l(k),k}$ is an affine or a Lipschits map. The following condition states formally the sampling requirements.

Condition SM. *The samples $D_k = \{(Y_{ik}, X_{ik}) : 1 \leq i \leq n_k\}$, $k \in \mathcal{K}$, are generated as follows: (a) For each population $k \in \mathcal{K}_0$, D_k contains i.i.d. copies of the random vector (Y_k, X_k) that has probability law P_k, and D_k are independent across $k \in \mathcal{K}_0$. (b) For each population $k \in \mathcal{K}_t$, the samples D_k are created by transformation, $D_k = \{g_{l(k),k}(Y_{il(k)}, X_{il(k)}) : 1 \leq i \leq n_{l(k)}\}$ for $l(k) \in \mathcal{K}_0$, where the maps $g_{l(k),k}$ preserve the Dudley-Pollard condition.*

Lemma E.4 in Appendix D shows the following result under Condition SM: As $n \to \infty$ the empirical processes $\widehat{\mathbb{G}}_k(f) := \frac{1}{\sqrt{n_k}} \sum_{i=1}^{n_k} f(Y_{ik}, X_{ik}) - \int f dP_k$ converge weakly,

$$\widehat{\mathbb{G}}_k(f) \rightsquigarrow \mathbb{G}_k(f), \tag{5.1}$$

as stochastic processes indexed by $(k, f) \in \mathcal{KF}$ in $\ell^\infty(\mathcal{KF})$. The limit processes $\mathbb{G}_k$ are tight P_k-Brownian bridges, which are independent across $k \in \mathcal{K}_0$,[17] and for $k \in \mathcal{K}_t$ are defined by:

$$\mathbb{G}_k(f) = \mathbb{G}_{l(k)}(f \circ g_{l(k),k}), \quad \forall f \in \mathcal{F}. \tag{5.2}$$

5.2. **Exchangeable bootstrap.** The following condition specifies how we should draw the bootstrap weights to preserve the dependence between the samples in the exchangeable bootstrap for estimators of counterfactual functionals described in Section 3.

Condition EB. *For each n_k and $k \in \mathcal{K}_0$, $(w_{k1}, ..., w_{kn_k})$ is an exchangeable,[18] nonnegative random vector, such that for some $\epsilon > 0$*

$$\sup_{n_k} E[w_{k1}^{2+\epsilon}] < \infty, \quad n_k^{-1} \sum_{i=1}^{n_k} (w_{ki} - \bar{w}_k)^2 \to_\mathbb{P} 1, \quad \bar{w}_k \to_\mathbb{P} 1 \geq 0, \tag{5.3}$$

where $\bar{w}_k = n_k^{-1} \sum_{i=1}^{n_k} w_{ki}$. Moreover, vectors $(w_{k1}, ..., w_{kn_k})$ are independent across $k \in \mathcal{K}_0$. For each $k \in \mathcal{K}_t$,

$$w_{ki} = w_{l(k)i}, \quad k \in \mathcal{K}_t. \tag{5.4}$$

---

[16]The definitions of suitably measurable and Pollard's entropy condition are recalled in Appendix A. Together with boundedness, these are well-known sufficient conditions for a function class to be universal Donsker (Dudley, 1987).

[17]A zero-mean Gaussian process $\mathbb{G}_k$ is a P_k-Brownian bridge if its covariance function takes the form $E[\mathbb{G}_k(f)\mathbb{G}_k(l)] = \int f l dP_k - \int f dP_k \int l dP_k$, for any f and l in $L^2(F_{X_k})$; see van der Vaart (1998).

[18]A sequence of random variables $X_1, X_2, ...$ is exchangeable if for any finite permutation σ of the indices $1, 2, ...$ the joint distribution of the permuted sequence $X_{\sigma(1)}, X_{\sigma(2)}, ...$ is the same as the joint distribution of the original sequence.

Remark 5.1 (Common bootstrap schemes). As pointed out in van der Vaart and Wellner (1996), by appropriately selecting the distribution of the weights, exchangeable bootstrap covers the most common bootstrap schemes as special cases. The empirical bootstrap corresponds to the case where $(w_{k1}, ..., w_{kn_k})$ is a multinomial vector with parameter n_k and probabilities $(1/n_k, ..., 1/n_k)$. The weighted bootstrap corresponds to the case where $w_{k1}, ..., w_{kn_k}$ are i.i.d. nonnegative random variables with $E[w_{k1}] = Var[w_{k1}] = 1$, e.g. standard exponential. The m out of n bootstrap corresponds to letting $(w_{k1}, ..., w_{kn_k})$ be equal to $\sqrt{n_k/m_k}$ times multinomial vectors with parameter m_k and probabilities $(1/n_k, ..., 1/n_k)$. The subsampling bootstrap corresponds to letting $(w_{k1}, ..., w_{kn_k})$ be a row in which the number $n_k(n_k - m_k)^{-1/2}m_k^{-1/2}$ appears m_k times and 0 appears $n_k - m_k$ times ordered at random, independent of the data. □

5.3. Inference theory for counterfactual estimators based on quantile regression. We proceed to impose the following standard conditions on (Y_j, X_j) for each $j \in \mathcal{J}$.

Condition QR. (a) The conditional quantile function takes the form $Q_{Y_j|X_j}(u|x) = x'\beta_j(u)$ for all $u \in \mathcal{U} = [\varepsilon, 1 - \varepsilon]$ with $0 < \varepsilon < 1/2$, and $x \in \mathcal{X}_j$. (b) The conditional density function $f_{Y_j|X_j}(y|x)$ exists, is uniformly continuous on (y, x) in the support of (Y_j, X_j), and is uniformly bounded. (c) The minimal eigenvalue of $J_j(u) = E[f_{Y_j|X_j}(X'_j\beta_j(u)|X_j)X_jX'_j]$ is bounded away from zero uniformly over $u \in \mathcal{U}$. (d) $E\|X_j\|^{2+\epsilon} < \infty$ for some $\epsilon > 0$.

In order to state the next result, let us define

$$\ell_{j,y,x}(Y_j, X_j) = f_{Y_j|X_j}(y|x)x'\psi_{j,F_{Y_j|X_j}(y|x)}(Y_j, X_j),$$

$$\psi_{j,u}(Y_j, X_j) = -J_j(u)^{-1}\{1(Y_j \leq X'_j\beta_j(u)) - u\}X_j,$$

$$\kappa_{jk,y}(Y_j, X_j, X_k) = \sqrt{s_j}\int \ell_{j,y,x}(Y_j, X_j)dF_{X_k}(x) + \sqrt{s_k}F_{Y_j|X_j}(y|X_k).$$

Theorem 5.1 (Validity of QR based counterfactual analysis). *Suppose that for each $j \in \mathcal{J}$, Conditions S, SM, and QR hold, the region of interest $\mathcal{Y}_j\mathcal{X}_j$ is a compact subset of $\mathbb{R}^{1+d_x}$, and $\mathcal{U}_j := \{u : x'\beta_j(u) \in \mathcal{Y}_j, \text{ for some } x \in \mathcal{X}_j\} \subseteq \mathcal{U}$. Then, (1) Condition D holds for the quantile regression estimator (3.7) of the conditional distribution and the empirical distribution estimator (3.4) of the covariate distribution. The limit processes are given by*

$$Z_j(y, x) = \mathbb{G}_j(\ell_{j,y,x}), \quad G_k(f) = \mathbb{G}_k(f), \quad (j, k) \in \mathcal{JK},$$

where $\mathbb{G}_k$ are the P_k-Brownian bridges defined in (5.1) and (5.2). In particular, $\{F_{Y_j|X_j}(y|\cdot) : y \in \mathcal{Y}_j\}$ is a universal Donsker class. (2) Exchangeable bootstrap consistently estimates the limit law of these processes under Condition EB. (3) Therefore, all conclusions of Theorems 4.1-4.2 and Corollaries 4.1 - 4.2 apply. In particular, the limit law for the estimated counterfactual distribution is given by $\bar{Z}_{jk}(y) := \mathbb{G}_j(\kappa_{jk,y})$, with covariance function $E[\bar{Z}_{jk}(y)\bar{Z}_{lm}(\bar{y})] = E[\kappa_{jk,y}\kappa_{lm,\bar{y}}] - E[\kappa_{jk,y}]E[\kappa_{lm,\bar{y}}]$.

This is the third main and new result of the paper. It derives the joint functional central limit theorem for the quantile regression estimator of the conditional distribution and the empirical

distribution function estimator of the covariate distribution. It also shows that exchangeable bootstrap consistently estimates the limit law. Moreover, the result characterizes the limit law of the estimator of the counterfactual distribution in Theorem 4.1, which in turn determines the limit laws of the estimators of the counterfactual quantile functions and other functionals, via Theorem 4.1 and Corollaries 4.1 and 4.2. Note that $\mathcal{U}_j \subseteq \mathcal{U}$ is the condition that permits the use of trimming in (3.7), since it says that the conditional distribution of Y_j given X_j on the region of interest $\mathcal{Y}_j \mathcal{X}_j$ is not determined by the tail conditional quantiles.

While proving Theorem 5.1, we establish the following corollary that may be of independent interest.

Corollary 5.1 (Validity of exchangeable bootstrap for QR coefficient process)**.** *Let* $\{(Y_{ji}, X_{ji}) : 1 \leq i \leq n_j\}$ *be a sample of i.i.d. copies of the random vector* (Y_j, X_j) *that has probability law* P_j *and obeys Condition QR. (1) As* $n_j \to \infty$*, the QR coefficient process possesses the following limit law:* $\sqrt{n_j}(\widehat{\beta}_j(\cdot) - \beta_j(\cdot)) \rightsquigarrow \mathbb{G}_j(\psi_{j,\cdot})$ *in* $\ell^\infty(\mathcal{U})^{d_x}$*, where* $\mathbb{G}_j$ *is a* P_j- *Brownian Bridge. (2) The exchangeable bootstrap law is consistent for the limit law, namely, as* $n_j \to \infty$*,*

$$\sqrt{n_j}(\widehat{\beta}_j^*(\cdot) - \widehat{\beta}_j(\cdot)) \rightsquigarrow_{\mathbb{P}} \mathbb{G}_j(\psi_{j,\cdot}) \ in \ \ell^\infty(\mathcal{U})^{d_x}.$$

The result (2) is new and shows that exchangeable bootstrap (which includes empirical bootstrap, weighted bootstrap, m out of n bootstrap, and subsampling) is valid for estimating the limit law of the entire QR coefficient process. Previously, such a result was available only for pointwise cases (e.g. Hahn, 1995, and Feng, He, and Hu, 2011), and the process result was available only for subsampling (Chernozhukov and Fernandez-Val, 2005, and Chernozhukov and Hansen, 2006).

5.4. **Inference Theory for Counterfactual Estimators based on Distribution Regression.** We shall impose the following condition on (Y_j, X_j) for each $j \in \mathcal{J}$.

Condition DR. *(a) The conditional distribution function takes the form* $F_{Y_j|X_j}(y|x) = \Lambda(x'\beta_j(y))$ *for all* $y \in \mathcal{Y}_j$ *and* $x \in \mathcal{X}_j$*, where* Λ *is either probit or logit link function. (b) The region of interest* $\mathcal{Y}_j$ *is either a compact interval in* $\mathbb{R}$ *or a finite subset of* $\mathbb{R}$*. In the former case, the conditional density function* $f_{Y_j|X_j}(y|x)$ *exists, is uniformly bounded and uniformly continuous in* (y,x) *in the support of* (Y_j, X_j)*. (c)* $E\|X_j\|^2 < \infty$ *and the minimum eigenvalue of*

$$J_j(y) := E\left[\frac{\lambda(X_j'\beta_j(y))^2}{\Lambda(X_j'\beta_j(y))[1 - \Lambda(X_j'\beta_j(y))]} X_j X_j'\right],$$

is bounded away from zero uniformly over $y \in \mathcal{Y}_j$*, where* λ *is the derivative of* Λ*.*

22

In order to state the next result, we define

$$\ell_{j,y,x}(Y_j, X_j) = \lambda(x'\beta_j(y))x'\psi_{j,y}(Y_j, X_j),$$

$$\psi_{j,y}(Y_j, X_j) = -J_j^{-1}(y)\frac{\Lambda(X_j'\beta_j(y)) - 1\{Y_j \le y\}}{\Lambda(X_j'\beta_j(y))(1 - \Lambda(X_j'\beta_j(y)))}\lambda(X_j'\beta_j(y))X_j,$$

$$\kappa_{jk,y}(Y_j, X_j, X_k) = \sqrt{s_j}\int \ell_{j,y,x}(Y_j, X_j)dF_{X_k}(x) + \sqrt{s_k}F_{Y_j|X_j}(y|X_k).$$

Theorem 5.2 (Validity of DR based counterfactual analysis). *Suppose that for each $j \in \mathcal{J}$, Conditions S, SM, and DR hold, and the region $\mathcal{Y}_j\mathcal{X}_j$ is a compact subset of $\mathbb{R}^{1+d_x}$. Then, (1) Condition D holds for the distribution regression estimator (3.5) of the conditional distribution and the empirical distribution estimator (3.4) of the covariate distribution, with limit processes given by*

$$Z_j(y,x) = \mathbb{G}_j(\ell_{j,y,x}), \quad G_k(f) = \mathbb{G}_k(f), \quad (j,k) \in \mathcal{JK},$$

where $\mathbb{G}_k$ are the P_k-Brownian bridges defined in (5.1) and (5.2). In particular, $\{F_{Y_j|X_j}(y|\cdot) : y \in \mathcal{Y}_j\}$ is a universal Donsker class. (2) Exchangeable bootstrap consistently estimates the limit law of these processes under Condition EB. (c) Therefore, all conclusions of Theorem 4.1 and 4.2, and of Corollaries 4.1 and 4.2 apply to this case. In particular, the limit law for the estimated counterfactual distribution is given by $\bar{Z}_{jk}(y) := \mathbb{G}_j(\kappa_{jk,y})$, with covariance function $E\bar{Z}_{jk}(y)\bar{Z}_{lm}(\bar{y}) = E[\kappa_{jk,y}\kappa_{lm,\bar{y}}] - E[\kappa_{jk,y}]E[\kappa_{lm,\bar{y}}].$

This is the fourth main and new result of the paper. It derives the joint functional central limit theorem for the distribution regression estimator of the conditional distribution and the empirical distribution function estimator of the covariate distribution. It also shows that bootstrap consistently estimates the limit law. Moreover, the result characterizes the limit law of the estimator of the counterfactual distribution in Theorem 4.1, which in turn determines the limit laws of the estimators of the counterfactual quantiles and other functionals, via Theorem 4.1 and Corollaries 4.1 and 4.2.

While proving Theorem 5.2, we also establish the following corollary that may be of independent interest.

Corollary 5.2 (Limit law and exchangeable bootstrap for DR coefficient process). *Let $\{(Y_{ji}, X_{ji}) : 1 \le i \le n_j\}$ be a sample of i.i.d. copies of the random vector (Y_j, X_j) that has probability law P_j and obeys Condition DR. (1) As $n_j \to \infty$, the DR coefficient process possesses the following limit law:*

$$\sqrt{n_j}(\widehat{\beta}_j(\cdot) - \beta_j(\cdot)) = \widehat{\mathbb{G}}_j(\psi_{j,.}) + o_{\mathbb{P}}(1) \rightsquigarrow \mathbb{G}_j(\psi_{j,.}) \text{ in } \ell^\infty(\mathcal{Y}_j)^{d_x},$$

where $\mathbb{G}_j$ is a P_j- Brownian Bridge. The exchangeable bootstrap law is consistent for the limit law, namely, as $n_j \to \infty$,

$$\sqrt{n_j}(\widehat{\beta}_j^*(\cdot) - \widehat{\beta}_j(\cdot)) \rightsquigarrow_{\mathbb{P}} \mathbb{G}_j(\psi_{j,.}) \text{ in } \ell^\infty(\mathcal{Y}_j)^{d_x}.$$

These limit distribution and bootstrap consistency results are new. They have already been applied in several studies (Chernozhukov, Fernandez-Val and Kowalski, 2011, Rothe, 2012, and Rothe and Wied, 2012). Note that unlike Theorem 5.2, this corollary does not rely on compactness of the region $\mathcal{Y}_j \mathcal{X}_j$.

6. Labor Market Institutions and the Distribution of Wages

In this section we apply our estimation and inference procedures to re-analyze the evolution of the U.S. wage distribution between 1979 and 1988. The first goal here is to compare the methods proposed in Section 3 and to discuss the various choices that practitioners need to make. The second goal is to provide support for the previous findings of DiNardo, Fortin, and Lemieux (1996, DFL hereafter) with a rigorous econometric analysis. Indeed, we provide confidence intervals for real-valued and function-valued effects of the institutional and labor market factors driving changes in the wage distribution, thereby quantifying their economic and statistical significance. We also provide a variance decomposition of the covariate composition effect into within-group and between-group components.

We use the same dataset and variables as in DFL, extracted from the outgoing rotation groups of the Current Population Surveys (CPS) in 1979 and 1988. The outcome variable of interest is the hourly log-wage in 1979 dollars. The regressors include a union status indicator, nine education dummy variables interacted with experience, a quartic term in experience, two occupation dummy variables, twenty industry dummy variables, and indicators for race, SMSA, marital status, and part-time status. Following DFL we weigh the observations by the product of the CPS sampling weights and the hours worked. We analyze the data only for men for the sake of brevity.[19]

The major factors suspected to have an important role in the evolution of the wage distribution between 1979 and 1988 are the minimum wage, whose real value declined by 27 percent, the level of unionization, whose level declined from 32 percent to 21 percent in our sample, and the characteristics of the labor force, whose education levels and other characteristics changed substantially during this period. Thus, following DFL, we decompose the total change in the US wage distribution into the sum of four effects: (1) the effect of a change in minimum wage, (2) the effect of de-unionization, (3) the effect of changes in the characteristics of the labor force other than unionization, and (4) the wage structure effect. We stress that this decomposition has a causal interpretation only under additional conditions analogous to the ones laid out in Section 2.3.

We formally define these four effects as differences between appropriately chosen counterfactual distributions. Let $F_{Y\langle(t,s)|(r,v)\rangle}$ denote the counterfactual distribution of log-wages Y when the wage structure is as in year t, the minimum wage M is at the level observed in year s, the union status U is distributed as in year r, and the other worker characteristics C are distributed as in year v. We use two indexes to refer to the conditional and covariate distributions because we treat

---

[19]Results for women can be found in Appendix C of the supplemental material.

the minimum wage as a feature of the conditional distribution and we want to separate union status from the other covariates. Given these counterfactual distributions, we can decompose the observed change in the distribution of wages between 1979 (year 0) and 1988 (year 1) into the sum of the previous four effects:

$$
\begin{aligned}
F_{Y\langle(1,1)|(1,1)\rangle} - F_{Y\langle(0,0)|(0,0)\rangle} \;=\; & \underbrace{[F_{Y\langle(1,1)|(1,1)\rangle} - F_{Y\langle(1,0)|(1,1)\rangle}]}_{(1)} + \underbrace{[F_{Y\langle(1,0)|(1,1)\rangle} - F_{Y\langle(1,0)|(0,1)\rangle}]}_{(2)} \\
+\; & \underbrace{[F_{Y\langle(1,0)|(0,1)\rangle} - F_{Y\langle(1,0)|(0,0)\rangle}]}_{(3)} + \underbrace{[F_{Y\langle(1,0)|(0,0)\rangle} - F_{Y\langle(0,0)|(0,0)\rangle}]}_{(4)}.
\end{aligned}
$$
(6.1)

In constructing the decompositions (6.1), we follow the same sequential order as in DFL.[20]

We next describe how to identify and estimate the various counterfactual distributions appearing in (6.1). The first counterfactual distribution is $F_{Y\langle(1,0)|(1,1)\rangle}$, the distribution of wages that we would observe in 1988 if the real minimum wage was as high as in 1979. Identifying this quantity requires additional assumptions.[21] Following DFL, the first strategy we employ is to assume the conditional wage density at or below the minimum wage depends only on the value of the minimum wage, and the minimum wage has no employment effects and no spillover effects on wages above its level. Under these conditions, DFL show that

$$
F_{Y_{(1,0)}|X_1}(y|x) = \begin{cases} F_{Y_{(0,0)}|X_0}(y|x)\, \dfrac{F_{Y_{(1,1)}|X_1}(m_0|x)}{F_{Y_{(0,0)}|X_0}(m_0|x)}, & \text{if } y < m_0; \\ F_{Y_{(1,1)}|X_1}(y|x), & \text{if } y \geq m_0; \end{cases}
$$
(6.2)

where $F_{Y_{(t,s)}|X_t}(y|x)$ denotes the conditional distribution of wages in year t given worker characteristics $X_t = (U_t, C_t)$ when the level of the minimum wage is as in year s, and m_s denotes the level of the minimum wage in year s. The second strategy we employ completely avoids modeling the conditional wage distribution below the minimal wage by simply censoring the observed wages below the minimum wage to the value of the minimum wage, i.e.

$$
F_{Y_{(1,0)}|X_1}(y|x) = \begin{cases} 0, & \text{if } y < m_0; \\ F_{Y_{(1,1)}|X_1}(y|x), & \text{if } y \geq m_0. \end{cases}
$$
(6.3)

Given either (6.2) or (6.3) we identify the counterfactual distribution of wages using the representation:

$$
F_{Y\langle(1,0)|(1,1)\rangle}(y) = \int F_{Y_{(1,0)}|X_1}(y|x)\,dF_{X_1}(x),
$$
(6.4)

where F_{X_t} is the joint distribution of worker characteristics and union status in year t. The other counterfactual marginal distributions we need are

$$
F_{Y\langle(1,0)|(0,1)\rangle}(y) = \int \int F_{Y_{(1,0)}|X_1}(y|x)\,dF_{U_0|C_0}(u|c)\,dF_{C_1}(c)
$$
(6.5)

---

[20]The sequential order may matter because it defines the counterfactual distributions and effects of interest. We report some results for the reverse sequential order in Appendix C of the supplemental material. The results are similar under the two alternative sequential orders.

[21]We cannot identify this quantity from random variation in minimum wage, since the same federal minimum wage applies to all individuals and state level minimum wages varied little across states in the years considered.

and

$$F_{Y\langle(1,0)|(0,0)\rangle}(y) = \int F_{Y_{(1,0)}|X_1}(y|x)\,dF_{X_0}(x)\,. \tag{6.6}$$

All the components of these distributions are identified and we estimate them using the plug-in principle. In particular, we estimate the conditional distribution $F_{U_0|C_0}(u|c), u \in \{0,1\}$, by logistic regression, and F_{X_1}, F_{C_1} and F_{X_0} by the empirical distributions.

From a practical standpoint, the main implementation decision concerns the choice of the estimator of the conditional distributions, $F_{Y_{(j,j)}|X_j}(y|x)$, for $j \in \{0,1\}$. We consider the use of quantile regression, distribution regression, classical regression, and duration/transformation regression. The classical regression and the duration regression models are parsimonious special cases of the first two models. However, these models are not appropriate in this application due to substantial conditional heteroskedasticity in log wages (Lemieux, 2006, and Angrist, Chernozhukov, and Fernandez-Val, 2006). As the additional restrictions that these two models impose are rejected by the data, we focus on the distribution and quantile regression approaches.

Distribution and quantile regressions impose different parametric restrictions on the data generating process. A linear model for the conditional quantile function may not provide a good approximation to the conditional quantiles near the minimum wage, where the conditional quantile function may be highly nonlinear. Indeed, under the assumptions of DFL the wage function has different determinants below and above the minimum wage. In contrast, the distribution regression model may well capture this type of behavior, since it allows the model coefficients to depend directly on the wage levels.

A second characteristic of our data set is the sizeable presence of mass points around the minimum wage and at some other round-dollar amounts. For instance, 20% of the wages take exactly 1 out of 6 values and 50% of the wages take exactly 1 out of 25 values. We compare the distribution and quantile regression estimators in a simulation exercise calibrated to fit many properties of the data set. The results presented in Appendix B of the supplemental material show that quantile regression is more accurate when the dependent variable is continuous but performs worse than distribution regression in the presence of realistic mass points. Based on these simulations and on specification tests that reject the linear quantile regression model, we employ the distribution regression approach to generate the main empirical results.[22] Since most of the problems for quantile regression take place in the region of the minimum wage, we also check the robustness of our results with a censoring approach. We censor wages from below at the value of the minimum wage and then apply censored quantile and distribution regressions to the resulting data.

We present our empirical results in Table 1 and Figures 1–3. In Table 1, we report the estimation and inference results for the decomposition (6.1) of the changes in various measures of wage dispersion between 1979 and 1988 estimated using logit distribution regressions. Figures

---

[22]Rothe and Wied (2012) propose new specification tests for conditional distribution models. Applying their tests to a similar dataset, they reject the quantile regression model but not the distribution regression model.

1-3 refine these results by presenting estimates and 95% simultaneous confidence intervals for several major counterfactual effects of interest, including quantile, distribution and Lorenz effects. We construct the simultaneous confidence bands using 100 bootstrap replications and a grid of quantile indices $\{0.02, 0.021, ..., 0.98\}$. We plot all of these function-valued effects against the quantile indices of wages.

We see in the top panels of Figures 1-3 that the low end of the distribution is significantly lower in 1988 while the upper end is significantly higher in 1988. This pattern reflects the well-known increase in wage inequality during this period. Next we turn to the decomposition of the total change into the sum of the four effects. For this decomposition we focus mostly on quantile functions for comparability with recent studies and to facilitate interpretation.[23] From Figure 1, we see that the contribution of de-unionization to the total change is quantitatively small and has a U-shaped effect across the quantile indexes. The magnitude and shape of this effect on the marginal quantiles between the first and last decile sharply contrast with the quantitatively large and monotonically decreasing shape of the effect of the union status on the conditional quantile function for this range of indexes (Chamberlain, 1994).[24] This comparison illustrates the difference between conditional and unconditional effects. The unconditional effects depend not only on the conditional effects but also on the characteristics of the workers who switched their unionization status. Obviously, de-unionization cannot affect those who were not unionized at the beginning of the period, which is 70 percent of the workers. In our data, the unionization rate declines from 32 to 21 percent, thus affecting only 11 percent of the workers. Thus, even though the conditional impact of switching from union to non-union status can be quantitatively large, it has a quantitatively small effect on the marginal distribution.

From Figure 1, we also see that the change in the distribution of worker characteristics (other than union status) is responsible for a large part of the increase in wage inequality. The importance of these composition effects has been recently stressed by Lemieux (2006) and Autor, Katz and Kearney (2008). The composition effect, including the de-unionization and worker characteristics effects, is realized through two channels: between-group and within-group inequality. To understand the effect of these channels on wage dispersion it is useful to consider a linear quantile model $Y = X'\beta(U)$, where X is independent of U. By the law of total variance, we can decompose the variance of Y into:

$$\text{Var}[Y] = E[\beta(U)]'\text{Var}[X]E[\beta(U)] + \text{trace}\{E[XX']\text{Var}[\beta(U)]\}, \tag{6.7}$$

---

[23]Discreteness of wage data implies that the quantile functions have jumps. To avoid this erratic behavior in the graphical representations of the results, we display smoothed quantile functions. The non-smoothed results are available from the authors. The quantile functions were smoothed using a bandwidth of 0.015 and a Gaussian kernel. The results in Table 1 have not been smoothed.

[24]We find similar estimates to Chamberlain (1994) for the effect of union status on the conditional quantile function in our CPS data.

where between-group inequality corresponds to the first term and within-group inequality corresponds to the second term.[25] When we keep the coefficients fixed, a change in the distribution of the covariates increases inequality through the first channel if the variance of the covariates increases and through the second channel if the proportion of high-variance groups increases. In our case, both components increased by about 10% between 1979 and 1988. The increase in the proportion of college graduate from 19% to 23% is an example of the observed composition changes. It raised between-group inequality because highly educated workers earn conditional average wages well above the unconditional average, and within-group inequality because of the higher wage volatility faced by these workers.[26]

We also include estimates of the wage structure effect, sometimes referred to as the price effect, which captures changes in the conditional distribution of log hourly wages. It represents the difference we would observe if the distribution of worker characteristics and union status, and the minimum wage remained unchanged during this period. This effect has a U-shaped pattern, which is similar to the pattern Autor, Katz and Kearney (2006a) find for the period between 1990 and 2000. They relate this pattern to a bi-polarization of employment into low and high skill jobs. However, they do not find a U-shaped pattern for the period between 1980 and 1990. A possible explanation for the apparent absence of this pattern in their analysis might be that the declining minimum wage masks this phenomenon. In our analysis, once we control for this temporary factor, we do uncover the U-shaped pattern for the price component in the 80s.

In Figure C1 of the supplemental material , we check the robustness of the results with respect to the link function used to implement the DR estimator. The results previously analyzed were obtained with a logistic link function. The differences between the estimates obtained with the logistic, normal, uniform (linear probability model), Cauchy and complementary log-log link functions are so modest that the lines are almost indistinguishable. As we mentioned above, the assumptions about the minimum wage are also delicate, since the mechanism that generates wages strictly below this level is not clear; it could be measurement error, non-coverage, or non-compliance with the law. To check the robustness of the results to the DFL assumptions about the minimum wage and to our semi-parametric model of the conditional distribution, we re-estimate the decomposition using censored linear quantile regression and censored distribution regression with a logit link, censoring the wage data below the minimum wage. For censored quantile regression, we use Powell's (1986) censored quantile regression estimated by Chernozhukov and Hong's (2002) algorithm. For censored distribution regression, we simply censor to zero the

---

[25]See Aaverge, Bjerve, and Doksum (2005) for an analogous decomposition of the pseudo-Lorenz curve. Similar within-between decompositions can also be constructed using distribution regression models.

[26]This is an empirical fact in our data set and not a theoretical fact. Increasing the proportion of educated workers can in principle either increase or decrease either component. To compute these effects, we kept the coefficients constant at their values obtained from estimating $F_{Y(1,0)|X}$ and changed the distribution of the covariates from dF_{X_0} to dF_{X_1}. See Appendix D of the supplemental material for more details on the computation of the variance decomposition.

distribution regression estimates of the conditional distributions below the minimum wage and recompute the functionals of interest. We find the results in Figure C2 to be very similar for the quantile and distribution regressions, and they are not very sensitive to the censoring.

Overall, our estimates and confidence intervals reinforce the findings of DFL, giving them a rigorous econometric foundation. Even though the sample size is large, the precision of some of the estimates was unclear to us a priori. For instance, only a relatively small proportion of workers are affected by unions. We provide standard errors and confidence intervals, which demonstrate the statistical and economic significance of the results. Moreover, we validate the results with a wide array of estimation methods. The similarity of the estimates may come as a surprise because the estimators make different parametric assumptions. However, in a fully saturated model all the estimators we have applied would give numerically the same results. The similarity of the results can be explained by the flexibility of our parametric model. Finally, we give a variance decomposition of the composition effect that shows that the increase in wage inequality is due to both between-group and within inequality components.

7. Conclusion and directions for future work

This paper develops methods for performing inference about the effect on an outcome of interest of a change in either the distribution of covariates or the relationship of the outcome with these covariates. The validity of the proposed inference procedures in large samples relies only on the applicability of a functional central limit theorem and the consistency of the bootstrap for the estimators of the covariate and conditional distributions. These conditions hold for the empirical distribution function estimator of the covariate distribution and for the most common regression estimators of the conditional distribution, such as classical, quantile, duration/transformation, and distribution regressions. Thus, we offer valid inference procedures for several popular existing estimators and introduce distribution regression to estimate counterfactual distributions.

We focus on functionals of the marginal counterfactual distributions but we do not consider their joint distribution. This joint distribution is required to compute other economically interesting quantities such as the distribution of the counterfactual effects. Abbring and Heckman (2007) discuss various ways to identify the distribution of these effects. The working paper version of this article provides inference procedures in one special case, rank invariance.

We focus on semi-parametric estimators of the conditional distribution due to their dominant role in empirical work (Angrist and Pischke, 2008). We hope to extend the analysis to nonparametric estimators in future work. Fully nonparametric estimators are in principle attractive but their implementation in samples of moderate size might be problematic. Rothe (2010) makes first steps in this direction and highlights some of the difficulties.

As mentioned in Foonote 1, our general results do not require the observability of the outcome of interest. If $F_{Y_j|X_j}(y|x)$ is redefined as the conditional distribution of a latent outcome and an estimator $\widehat{F}_{Y_j|X_j}(y|x)$ that satisfies Condition D is available, then the inference results in Section

29

4 apply. An interesting example is given by the policy relevant treatment effects of Heckman and Vytlacil (2005). They consider a class of policies that affect the probability of participation in a program but do not affect directly the structural function of the outcome in a model with endogeneity. For instance, one may be interested in the effect of decreasing college tuition on wages. In their model, the policy relevant treatment effect is the conditional marginal treatment effect integrated over the covariate distribution and the error term in the participation equation. This type of policy effects is outside of the scope of this paper but is certainly worth pursuing in future research.

Appendix A. Notation

Given a weakly increasing function $F : \mathcal{Y} \subseteq \mathbb{R} \mapsto \mathcal{T} \subseteq [0,1]$, we define the left-inverse of F as the function $F^{\leftarrow} : \mathcal{T} \mapsto \overline{\mathcal{Y}}$, where $\overline{\mathcal{Y}}$ is the closure of $\mathcal{Y}$, such that

$$
F^{\leftarrow}(\tau) = \begin{cases} \inf\{y \in \mathcal{Y} : F(y) \geq \tau\} & \text{if } \sup_{y \in \mathcal{Y}} F(y) > \tau, \\ \sup\{y \in \mathcal{Y}\} & \text{otherwise.} \end{cases}
$$

Each sample for the population k is defined on a probability space $(\Omega_k, \mathcal{A}_k, P_k)$, and there is an underlying common probability space $(\Omega, \mathcal{A}, \mathbb{P})$ that contains the product $\times_{k \in \mathcal{K}}(\Omega_k, \mathcal{A}_k, P_k)$. We write $Z_n \rightsquigarrow Z$ in $\mathbb{E}$ to denote the weak convergence of a stochastic process Z_n to a random element Z in a normed space $\mathbb{E}$, as defined in van der Vaart and Wellner (1996) (VW). We write $\rightarrow_{\mathbb{P}}$ to denote convergence in outer probability. We write $\rightsquigarrow_{\mathbb{P}}$ to denote the weak convergence of the bootstrap law in probability, as formally defined in Section 4. Given the sequences of stochastic processes $Z_{m1}, ..., Z_{mn}, m \in \mathcal{M}$ for some finite set $\mathcal{M}$, taking values in normed spaces $\mathbb{E}_m$, we say that $Z_{mn} \rightsquigarrow Z_m$ jointly in $m \in \mathcal{M}$, if $(Z_{mn} : m \in \mathcal{M}) \rightsquigarrow (Z_m : m \in \mathcal{M})$ in $\mathbb{E} = \times_{m \in \mathcal{M}} \mathbb{E}_m$, where the product space $\mathbb{E}$ is endowed with the norm $\|\cdot\|_{\mathbb{E}} = \vee_{m \in \mathcal{M}} \|\cdot\|_{\mathbb{E}_m}$, see Section 1.4 in VW. The space $\ell^{\infty}(\mathcal{F})$ represents the space of real-valued bounded functions defined on the index set equipped with the supremum norm $\|\cdot\|_{\ell^{\infty}(\mathcal{F})}$. Following VW, we use the simplified notation $\|\cdot\|_{\mathcal{F}}$ to denote the supremum norm. Given a measurable subset $\mathcal{X}$ of $\mathbb{R}^k$, a class $\mathcal{F}$ of measurable functions $f : \mathcal{X} \to \mathbb{R}$ is called a universal Donsker class if for every probability measure P on $\mathcal{X}$, $\sqrt{n}(P_n - P) \rightsquigarrow \mathbb{G}$ in $\ell^{\infty}(\mathcal{F})$, where P_n is the empirical measure and $\mathbb{G}$ is a P-Brownian bridge (Dudley, 1987). By Dudley (1987) a sufficient condition for a uniformly bounded class of measurable functions $\mathcal{F}$ to be universal Donsker is the Pollard's entropy condition, which requires the uniform covering entropy integral for $\mathcal{F}$ to be finite, and suitable mesurability, namely that $\mathcal{F}$ is an image admissible Suslin class (Dudley, 1987). We call these conditions the Dudley-Pollard condition, and call a class of functions $\mathcal{F}$ that obeys them a Dudley-Pollard class. The measurability condition, developed by Dudley (1987), is mild and holds in most applications, including in our analysis. We do not explicitly discuss this condition in what follows.

We shall use the functional delta method, as formulated in VW. Let $\mathbb{D}_0$, $\mathbb{D}$, and $\mathbb{E}$ be normed spaces, with $\mathbb{D}_0 \subset \mathbb{D}$. A map $\phi : \mathbb{D}_\phi \subset \mathbb{D} \mapsto \mathbb{E}$ is called *Hadamard-differentiable* at $\theta \in \mathbb{D}_\phi$ tangentially to $\mathbb{D}_0$ if there is a continuous linear map $\phi'_\theta : \mathbb{D}_0 \mapsto \mathbb{E}$ such that

$$\frac{\phi(\theta + t_n h_n) - \phi(\theta)}{t_n} \to \phi'_\theta(h), \quad n \to \infty,$$

for all sequences $t_n \to 0$ and $h_n \to h \in \mathbb{D}_0$ such that $\theta + t_n h_n \in \mathbb{D}_\phi$ for every n.

Lemma B.1 (Functional delta-method). *Let $\mathbb{D}_0$, $\mathbb{D}$, and $\mathbb{E}$ be normed spaces. Let $\phi : \mathbb{D}_\phi \subset \mathbb{D} \mapsto \mathbb{E}$ be Hadamard-differentiable at θ tangentially to $\mathbb{D}_0$. Let $X_n : \Omega_n \mapsto \mathbb{D}_\phi$ be maps with $r_n(X_n - \theta) \rightsquigarrow X$ in $\mathbb{D}$, where X is separable and takes its values in $\mathbb{D}_0$, for some sequence of constants $r_n \to \infty$. Then $r_n(\phi(X_n) - \phi(\theta)) \rightsquigarrow \phi'_\theta(X)$. If ϕ'_θ is defined and continuous on the whole of $\mathbb{D}$, then the sequence $r_n(\phi(X_n) - \phi(\theta)) - \phi'_\theta(r_n(X_n - \theta))$ converges to zero in outer probability.*

The applicability of the method is greatly enhanced by the fact that Hadamard differentiation obeys the chain rule, for a formal statement of which we refer to VW. We also use the following simple "stacking rule" in the proofs.

Lemma B.2 (Stacking rule). *If $\phi_1 : \mathbb{D}_{\phi_1} \subset \mathbb{D}_1 \mapsto \mathbb{E}_1$ is Hadamard-differentiable at $\theta_1 \in \mathbb{D}_{\phi_1}$ tangentially to $\mathbb{D}_{10}$ with derivative $\phi'_{1\theta_1}$ and $\phi_2 : \mathbb{D}_{\phi_2} \subset \mathbb{D}_2 \mapsto \mathbb{E}_2$ is Hadamard-differentiable at $\theta_2 \in \mathbb{D}_{\phi_2}$ tangentially to $\mathbb{D}_{20}$ with derivative $\phi'_{2\theta_2}$, then $\phi = (\phi_1, \phi_2) : \mathbb{D}_{\phi_1} \times \mathbb{D}_{\phi_2} \subset \mathbb{D}_1 \times \mathbb{D}_2 \mapsto \mathbb{E}_1 \times \mathbb{E}_2$ is Hadamard-differentiable at $\theta = (\theta_1, \theta_2)$ tangentially to $\mathbb{D}_{01} \times \mathbb{D}_{02}$ with derivative $\phi'_\theta = (\phi'_{1\theta_1}, \phi'_{2\theta_2})$.*

Let D_n denote the data vector and M_n be a vector of random variables, used to generate bootstrap draws or simulation draws (this may depend on particular method). Consider sequences of random elements $V_n = V_n(D_n)$ and $G_n^* = G_n(D_n, M_n)$ in a normed space $\mathbb{D}$, where the sequence $G_n = \sqrt{n}(V_n - V)$ weakly converges unconditionally to the tight random element G, and G_n^* converges conditionally given D_n in distribution to G, in probability, denoted as $G_n \rightsquigarrow G$ and $G_n^* \rightsquigarrow_\mathbb{P} G$, respectively.[27] Let $V_n^* = V_n + G_n^*/\sqrt{n}$ denote the bootstrap or simulation draw of V_n.

Lemma B.3 (Delta-method for bootstrap and other simulation methods). *Let $\mathbb{D}_0$, $\mathbb{D}$, and $\mathbb{E}$ be normed spaces, with $\mathbb{D}_0 \subset \mathbb{D}$. Let $\phi : \mathbb{D}_\phi \subset \mathbb{D} \mapsto \mathbb{E}$ be Hadamard-differentiable at V tangentially to $\mathbb{D}_0$. Let V_n and V_n^* be maps as indicated previously with values in $\mathbb{D}_\phi$ such that $\sqrt{n}(V_n - V) \rightsquigarrow G$ and $\sqrt{n}(V_n^* - V_n) \rightsquigarrow_\mathbb{P} G$, where G is separable and takes its values in $\mathbb{D}_0$. Then in $\mathbb{E}$ $\sqrt{n}(\phi(V_n^*) - \phi(V_n)) \rightsquigarrow_\mathbb{P} \phi'_V(G)$.*

Another technical result that we use in the sequel concerns the equivalence of continuous and uniform convergence.

---

[27]This standard concept is recalled in Section 4; see also VW, Chap. 3.9.

Lemma B.4 (Uniform convergence via continuous convergence)**.** *Let $\mathbb{D}$ and $\mathbb{E}$ be complete separable metric spaces, with $\mathbb{D}$ compact. Suppose $f : \mathbb{D} \mapsto \mathbb{E}$ is continuous. Then a sequence of functions $f_n : \mathbb{D} \mapsto \mathbb{E}$ converges to f uniformly on $\mathbb{D}$ if and only if for any convergent sequence $x_n \to x$ in $\mathbb{D}$ we have that $f_n(x_n) \to f(x)$.*

For the proofs of Lemmas B.1 and B.3, see VW, Chap. 1.11 and 3.9. Lemma B.2 follows from the definition of Hadamard derivative and product space. For the proof of Lemma B.4, see, for example, Resnick (1987), page 2.

<div align="center">APPENDIX C. PROOF OF LEMMA 2.1</div>

First, note that $Y = \sum_{j \in \mathcal{J}} 1(J = j) Y_j^*$, so that

$$F_{Y|J,X}(y \mid j, x) = F_{Y_j^*|J,X}(y \mid j, x).$$

Also, $Y_j = Y|J = j$ and $X_k = X|J = k$, so that $F_{Y|J,X}(y \mid j, x) \equiv F_{Y_j|X_j}(y \mid x)$ and $F_{X|J}(x \mid k) \equiv F_{X_k}(x)$. Hence, by iterating expectations

$$
\begin{aligned}
F_{Y_j^*|J}(y \mid k) = \int_{\mathcal{X}_k} F_{Y_j^*|J,X}(y \mid k, x) dF_{X|J}(x \mid k) &= \int_{\mathcal{X}_k} F_{Y_j^*|J,X}(y \mid j, x) dF_{X|J}(x \mid k) \\
&= \int_{\mathcal{X}_k} F_{Y_j|X_j}(y \mid x) dF_{X_k}(x),
\end{aligned}
$$

where the second equality follows by conditional exogeneity (2.8), and the last uses the facts stated above. $\qquad\square$

<div align="center">APPENDIX D. PROOF OF THEOREMS 4.1–4.2 AND COROLLARIES 4.1–4.2.</div>

D.1. **Key ingredient: Hadamard differentiability of counterfactual operator.** It suffices to consider a single pair $(j, k) \in \mathcal{JK}$. In order to keep the notation simple, we drop the indices (j, k) wherever possible.

We need some setup and preliminary observations. Let $\ell_m^\infty(\mathcal{YX})$ denote the set of all bounded and measurable mappings $\mathcal{YX} \mapsto \mathbb{R}$. Let $\mathcal{F}$, Z, and G be specified as in Condition D, with the indices (j, k) omitted from the subscripts. We consider $\mathcal{YX}$ as a subset of $\overline{\mathbb{R}}^{1+d_x}$, with relative topology. Let ρ denote a standard metric on $\overline{\mathbb{R}}^{1+d_x}$. The closure of $\mathcal{YX}$ under ρ, denoted $\overline{\mathcal{YX}}$, is compact in $\overline{\mathbb{R}}^{1+d_x}$. By Condition D, a.s. Z takes values in $UC(\mathcal{YX}, \rho)$, the set of functions mapping $\mathcal{YX}$ to the real line that are uniformly continuous with respect to metric ρ, and can be continuously extended to $\overline{\mathcal{YX}}$, so that $UC(\mathcal{YX}, \rho) \subset \ell_m^\infty(\mathcal{YX})$. By Condition D, $G \in UC(\mathcal{F}, \lambda)$ a.s., where $\lambda(f, \tilde{f}) = [P(f - \tilde{f})^2]^{1/2}$ is a (semi) metric on $\mathcal{F}$.

Lemma D.1 (Hadamard differentiability of counterfactual operator)**.** *Let $\mathcal{YX} \subseteq \mathbb{R}^{1+d_x}$, and $\mathcal{F}$ be the class of bounded functions, mapping $\overline{\mathbb{R}}^{d_x}$ to $\mathbb{R}$, that contains $\{F_{Y|X}(y|\cdot) : y \in \mathcal{Y}\}$ as well as the indicators of all the rectangles in $\overline{\mathbb{R}}^{d_x}$. Let $\mathbb{D}_\phi$ be the product of the space of measurable functions $\Gamma : \mathcal{YX} \mapsto [0, 1]$ defined by $(y, x) \mapsto \Gamma(y, x)$ and the bounded maps $\Pi : \mathcal{F} \mapsto \mathbb{R}$ defined*

<div align="center">32</div>

by $f \mapsto \int f d\Pi$, where Π is restricted to be a probability measure on $\mathcal{X}$. Consider the map $\phi : \mathbb{D}_\phi \subset \mathbb{D} = \ell_m^\infty(\mathcal{YX}) \times \ell^\infty(\mathcal{F}) \mapsto \mathbb{E} = \ell^\infty(\mathcal{Y})$, defined by

$$(\Gamma, \Pi) \mapsto \phi(\Gamma, \Pi) := \int \Gamma(\cdot, x) d\Pi(x).$$

Then the map ϕ is well defined. Moreover, the map ϕ is Hadamard-differentiable at $(\Gamma, \Pi) = (F_{Y|X}, F_X)$, tangentially to the subset $\mathbb{D}_0 = UC(\mathcal{YX}, \rho) \times UC(\mathcal{F}, \lambda)$, with the derivative map $(\gamma, \pi) \mapsto \phi'_{F_{Y|X}, F_X}(\gamma, \pi)$ mapping $\mathbb{D}$ to $\mathbb{E}$ defined by

$$\phi'_{F_{Y|X}, F_X}(\gamma, \pi)(y) := \int \gamma(y, x) dF_X(x) + \pi(F_{Y|X}(y|\cdot)),$$

where the derivative is defined and is continuous on $\mathbb{D}$.

Proof of Lemma D.1. First we show that the map is well defined. Any probability measure Π on $\mathcal{X}$ is determined by the values $\int f d\Pi$ for $f \in \mathcal{F}$, since $\mathcal{F}$ contains all the indicators of the rectangles in $\mathbb{R}^{d_x}$. By Caratheodory's extension theorem $\Pi(A) = \Pi 1_A$ is well defined on all Borel subsets A of $\mathbb{R}^{d_x}$. Since $x \mapsto \Gamma(y, x)$ is Borel measurable and takes values in $[0, 1]$, it follows that $\int \Gamma(y, x) d\Pi(x)$ is well defined as a Lebesgue integral, and $\int \Gamma(\cdot, x) d\Pi(x) \in \ell^\infty(\mathcal{Y})$.

Next we show the main claim. Consider any sequence $(\Gamma^t, \Pi^t) \in \mathbb{D}_\phi$ such that for $\gamma^t := (\Gamma^t - F_{Y|X})/t$, and $\pi^t(f) := \int f d(\Pi^t - F_X)/t$,

$$(\gamma^t, \pi^t) \to (\gamma, \pi), \quad \text{in} \quad \ell_m^\infty(\mathcal{YX}) \times \ell^\infty(\mathcal{F}), \quad \text{where } (\gamma, \pi) \in \mathbb{D}_0.$$

We want to show that as $t \searrow 0$

$$\frac{\phi(\Gamma^t, \Pi^t) - \phi(F_{Y|X}, F_X)}{t} - \phi'_{F_{Y|X}, F_X}(\gamma, \pi) \to 0 \text{ in } \ell^\infty(\mathcal{Y}).$$

Write the difference above as

$$\int (\gamma^t(y, x) - \gamma(y, x)) dF_X(x) + (\pi^t - \pi)(F_{Y|X}(y|\cdot)) + t\pi^t(\gamma(y|\cdot)) + t\pi^t(\gamma^t(y|\cdot) - \gamma(y|\cdot))$$

$$=: i(y) + ii(y) + iii(y) + iv(y).$$

Since $\gamma^t \to \gamma$ in $\ell_m^\infty(\mathcal{YX})$, we have that $\|i\|_{\mathcal{Y}} \leq \|\gamma^t - \gamma\|_{\mathcal{YX}} \int dF_X \to 0$, where $\|\cdot\|_{\mathcal{YX}}$ is the supremum norm in $\ell_m^\infty(\mathcal{YX})$ and $\|\cdot\|_{\mathcal{Y}}$ is the supremum norm in $\ell^\infty(\mathcal{Y})$. Moreover, since $\pi^t \to \pi$ in $\ell^\infty(\mathcal{F})$ and $\{F_{Y|X}(y|\cdot) : y \in \mathcal{Y}\} \subset \mathcal{F}$ by assumption, we have $\|ii\|_{\mathcal{Y}} \leq \|\pi^t - \pi\|_{\mathcal{F}} \to 0$, where $\|\cdot\|_{\mathcal{F}}$ is the supremum norm in $\ell^\infty(\mathcal{F})$. Further,

$$\|iv\|_{\mathcal{Y}} = \left\| \int (\gamma^t - \gamma)(\cdot, x) dt\pi^t(x) \right\|_{\mathcal{Y}} \leq \|\gamma^t - \gamma\|_{\mathcal{YX}} \int |d(\Pi^t - F_X)| \leq \|\gamma^t - \gamma\|_{\mathcal{YX}} \, 2 \to 0,$$

since $td\pi^t = d(\Pi^t - F_X)$ and $\int |d(\Pi^t - F_X)| \leq \int d\Pi^t + \int dF_X \leq 2$, where $\int |d\mu|$ indicates the total variation of a signed measure μ.

Since γ is continuous on the compact semi-metric space $(\overline{\mathcal{YX}}, \rho)$, there exists a finite partition of $\overline{\mathbb{R}}^{1+d_x}$ into non-overlapping rectangular regions $(R_{im} : 1 \leq i \leq m)$ (rectangles are allowed not to include their sides to make them non-overlapping) such that γ varies at most ϵ on $\mathcal{YX} \cap R_{im}$.

33

Let $p_m(y,x) := (y_{im}, x_{im})$ if $(y,x) \in \mathcal{YX} \cap R_{im}$, where (y_{im}, x_{im}) is an arbitrarily chosen point within $\mathcal{YX}_{im}$ for each i; also let $\chi_{im}(y,x) := 1\{(y,x) \in R_{im}\}$. Then, as $t \to 0$,

$$\|iii\|_{\mathcal{Y}} = \left\| \int \gamma(\cdot, x) t d\pi^t(x) \right\|_{\mathcal{Y}} \leq \left\| \int (\gamma - \gamma \circ p_m)(\cdot, x) t d\pi^t(x) \right\|_{\mathcal{Y}} + \left\| \int (\gamma \circ p_m)(\cdot, x) t d\pi^t(x) \right\|_{\mathcal{Y}}$$

$$\leq \|\gamma - \gamma \circ p_m\|_{\mathcal{YX}} \int |t d\pi^t| + \sum_{i=1}^{m} |\gamma(y_{im}, x_{im})| t \pi^t(\chi_{im})$$

$$\leq \|\gamma - \gamma \circ p_m\|_{\mathcal{YX}} 2 + \sum_{i=1}^{m} |\gamma(y_{im}, x_{im})| t \pi^t(\chi_{im}) \leq 2\epsilon + \sum_{i=1}^{m} |\gamma(y_{im}, x_{im})| t (\pi(\chi_{im}) + o(1))$$

$$\leq 2\epsilon + tm \left[\|\gamma\|_{\mathcal{YX}} \|\pi\|_{\mathcal{F}} + o(1) \right] \leq 2\epsilon + O(t) \to 2\epsilon,$$

since $\{\chi_{im} : 1 \leq i \leq m\} \subset \mathcal{F}$, so that $\pi^t(\chi_{im}) \to \pi(\chi_{im}) \leq \|\pi\|_{\mathcal{F}} < \infty$, for all $1 \leq i \leq m$, where $\|\cdot\|_{\mathcal{F}}$ denotes the supremum norm in $\ell^\infty(\mathcal{F})$. The constant ϵ is arbitrary, so $\|iii\|_{\mathcal{Y}} \to 0$ as $t \to 0$.

Note that the derivative is well-defined over the entire $\mathbb{D}$ and is in fact continuous with respect to the norm on $\mathbb{D}$ given by $\|\cdot\|_{\mathcal{YX}} \vee \|\cdot\|_{\mathcal{F}}$. The second component of the derivative map is trivially continuous with respect to $\|\cdot\|_{\mathcal{F}}$. The first component is continuous with respect to $\|\cdot\|_{\mathcal{YX}}$ since

$$\left\| \int (\gamma(\cdot, x) - \tilde{\gamma}(\cdot, x)) dF_X(x) \right\|_{\mathcal{Y}} \leq \|\gamma - \tilde{\gamma}\|_{\mathcal{YX}} \int dF_X(x).$$

Hence the derivative map is continuous. □

D.2. **Proof of Theorems 4.1 and 4.2.** In the notation of Lemma D.1, $\widehat{F}_{Y\langle j|k\rangle}(\cdot) = \phi(\widehat{F}_{Y_j|X_j}, \widehat{F}_{X_k})(\cdot) = \int \widehat{F}_{Y_j|X_j}(\cdot|x) d\widehat{F}_{X_k}(x)$ and $F_{Y\langle j|k\rangle}(\cdot) = \phi(F_{Y_j|X_j}, F_{X_k})(\cdot) = \int F_{Y_j|X_j}(\cdot|x) dF_{X_k}(x)$. The main result needed to prove the theorem is provided by Lemma D.1, which established that the map ϕ is Hadamard differentiable. This result holds uniformly in $(j,k) \in \mathcal{JK}$, since $\mathcal{JK}$ is a finite set. Moreover, under condition S, condition D can be restated as:

$$\left(\sqrt{n}(\widehat{F}_{Y_j|X_j}(y|x) - F_{Y_j|X_j}(y|x)), \sqrt{n} \int f d(\widehat{F}_{X_k} - F_{X_k}) \right) \rightsquigarrow \left(\sqrt{s_j} Z_j(y,x), \sqrt{s_k} G_k(f) \right),$$

as stochastic processes indexed by $(y,x,j,k,f) \in \mathcal{YXJKF}$ in the metric space $\ell^\infty(\mathcal{YXJKF})^2$. By the Functional Delta Method quoted in Lemma B.1, it follows that

$$\sqrt{n}(\widehat{F}_{Y\langle j|k\rangle} - F_{Y\langle j|k\rangle})(y) = \sqrt{s_j} \int \sqrt{n}[\widehat{F}_{Y_j|X_j}(y|x) - F_{Y_j|X_j}(y|x)] dF_{X_k}(x)$$

$$+ \sqrt{s_k} \int F_{Y_j|X_j}(y|x) \sqrt{n} d[\widehat{F}_{X_k}(x) - F_{X_k}(x)] + o_{\mathbb{P}}(1) \qquad (\text{D.1})$$

$$\rightsquigarrow \bar{Z}_{jk}(y) := \sqrt{s_j} \int Z_j(y,x) dF_{X_k}(x) + \sqrt{s_k} G_k(F_{Y_j|X_j}(y|\cdot)),$$

jointly in $(j,k) \in \mathcal{JK}$. The first order expansion (D.1) above is not needed to prove the theorem, but it can be useful for other applications. The continuity of the sample paths of $\bar{Z}_{jk}$ follows from the continuity of the sample paths of $Z_j(y,x)$ with respect to (y,x) and from the continuity of the sample paths of $G_k(f)$ with respect to f under the metric λ, by Condition D. Mean square

34

continuity of $F_{Y_j|X_j}(y|\cdot)$ with respect to y therefore implies continuity of the sample paths of $y \mapsto G_k(F_{Y_j|X_j}(y|\cdot))$. The first claim thus is proven.

In order to show the second claim, we first examine in detail the simple case where $y \mapsto \widehat{F}_{Y\langle j|k\rangle}(y)$ is weakly increasing in y. (For example, qr-based estimators are necessarily weakly increasing, while dr-based estimators need not be.) In this case $\widehat{Q}_{Y\langle j|k\rangle} = \widehat{F}^{\leftarrow}_{Y\langle j|k\rangle}$ and Hadamard differentiability of quantile (left inverse) operator (Doss and Gill, 1992, VW) implies by the Functional Delta Method:

$$\sqrt{n}\left(\widehat{Q}_{Y\langle j|k\rangle}(\tau) - Q_{Y\langle j|k\rangle}(\tau)\right) = -\frac{\sqrt{n}(\widehat{F}_{Y\langle j|k\rangle} - F_{Y\langle j|k\rangle})}{f_{Y\langle j|k\rangle}}(Q_{Y\langle j|k\rangle}(\tau)) + o_{\mathbb{P}}(1) \qquad (\text{D.2})$$

$$\rightsquigarrow \frac{\bar{Z}_{jk}}{f_{Y\langle j|k\rangle}}(Q_{Y\langle j|k\rangle}(\tau)), \qquad (\text{D.3})$$

as a stochastic process indexed by $(\tau, j, k) \in \mathcal{TJK}$ in the metric space $\ell^\infty(\mathcal{TJK})$.

When $y \mapsto \widehat{F}_{Y\langle j|k\rangle}(y)$ is not weakly increasing, the previous argument does not apply because the references cited above require $\widehat{F}_{Y\langle j|k\rangle}$ to be a proper distribution function. In this case, with probability converging to one we have that $\widehat{Q}_{Y\langle j|k\rangle} := \widehat{F}^{r\leftarrow}_{Y\langle j|k\rangle}$, where $\widehat{F}^r_{Y\langle j|k\rangle}$ is rearrangement of $\widehat{F}_{Y\langle j|k\rangle}$ on the interval $[a, b]$ defined in the statement of the theorem. In order to establish the properties of this estimator, we first recall the relevant result on Hadamard differentiability of the monotone rearrangement operator derived by Chernozhukov, Fernandez-Val, and Galichon (2010). Let F be a continuously differentiable function on the interval $[a, b]$ with strictly positive derivative f. Consider the rearrangement map $G \mapsto G^r$, which maps bounded measurable functions G on the domain $[a, b]$ and produces cadlag functions G^r on the same domain. This map, considered as a map $\ell^\infty_m([a, b]) \mapsto \ell^\infty_m([a, b])$, is Hadamard differentiable at F tangentially to $C([a, b])$, with the derivative map given by the identity $g \mapsto g$ which is defined and continuous on the whole $\ell^\infty_m([a, b])$. Therefore, we conclude by the Functional Delta Method that for all $(j, k) \in \mathcal{JK}$, $\sqrt{n}(\widehat{F}^r_{Y\langle j|k\rangle} - F_{Y\langle j|k\rangle})(\cdot) = \sqrt{n}(\widehat{F}_{Y\langle j|k\rangle} - F_{Y\langle j|k\rangle})(\cdot) + o_{\mathbb{P}}(1)$. Hence the rearranged estimator is first order equivalent to the original estimator and thus inherits the limit distribution. Now apply the differentiability of the quantile operator and the delta method again to reach the same final conclusions (D.2)- (D.3) as above.

Theorem 4.2 follows from the application of the functional delta method for the (generalized) bootstrap quoted in Lemma B.3 and the chain rule for the Hadamard derivative. $\qquad \square$

D.3. **Proof of Corollaries 4.1–4.2.** Corollary 4.1 follows from Theorem 4.1 by the Extended Continuous Mapping theorem. Corollary 4.2 follows by the Functional Delta Method. $\qquad \square$

Appendix E. Proof of Theorem 5.1 and 5.2

It is convenient to organize the proof in several steps. The task is complex: We need to show convergence and bootstrap convergence simultaneously for estimators of conditional distributions based on QR or DR and of estimators of covariate distributions based on empirical measures.

Since both distribution and quantile regression processes are Z-processes, we can complete the task efficiently by using Hadamard differentiability of the so called Z-maps. Hence in Section E.1 we present a functional delta method for Z-maps (Lemma E.2) and show how to apply it to a generic Z-problem (Lemma E.3). The results of this section are of independent interest. In Section E.2 we present the proofs for Section E.1. In Section E.3 we present the results on convergence of empirical measures, which take into account dependencies across samples in the presence of transformation samples. Finally, with all these ingredients, we prove Theorems 5.1 and 5.2 in Sections E.4 and E.5.

E.1. **Main ingredient: functional delta method for Z-processes.** In our leading examples, we have a functional parameter p-vector $u \mapsto \theta(u)$ where $u \in \mathcal{U}$ and $\theta(u) \in \Theta \subseteq \mathbb{R}^p$, and, for each $u \in \mathcal{U}$, the value $\theta_0(u)$ solves the p-vector of moment equations $\Psi(\theta, u) = 0$. For estimation purposes we have an empirical analog of the above moment functions $\widehat{\Psi}(\theta, u)$. For each $u \in \mathcal{U}$, the estimator $\widehat{\theta}(u)$ satisfies

$$\|\widehat{\Psi}(\widehat{\theta}(u), u)\|^2 \leq \inf_{\theta \in \Theta} \|\widehat{\Psi}(\theta, u)\|^2 + \widehat{r}(u)^2,$$

with $\|\widehat{r}\|_{\mathcal{U}} = o_{\mathbb{P}}(n^{-1/2})$. Similarly suppose that a bootstrap or simulation method is available that produces a pair $(\widehat{\Psi}^*, \widehat{r}^*)$ and the corresponding estimator $\widehat{\theta}^*(u)$ that obeys $\|\widehat{\Psi}^*(\widehat{\theta}^*(u), u)\|^2 \leq \inf_{\theta \in \Theta} \|\widehat{\Psi}^*(\theta, u)\|^2 + \widehat{r}^*(u)^2$, with $\|\widehat{r}^*\|_{\mathcal{U}} = o_{\mathbb{P}}(n^{-1/2})$.

We can represent the above estimator and estimand as

$$\widehat{\theta}(\cdot) = \phi(\widehat{\Psi}(\cdot, \cdot), \widehat{r}(\cdot)) \text{ and } \theta_0(\cdot) = \phi(\Psi(\cdot, \cdot), 0)$$

where ϕ is a *Z-map* formally defined as follows. Consider a p-vector $\psi(\theta, u)$ indexed by (θ, u) as a generic value of Ψ. An element $\theta \in \Theta$ is an $r(u)$-approximate zero of the map $\theta \mapsto \psi(\theta, u)$ if

$$\|\psi(\theta, u)\|^2 \leq \inf_{\theta' \in \Theta} \|\psi(\theta', u)\|^2 + r(u)^2,$$

where $r(u) \in \mathbb{R}$ is a numerical tolerance parameter. Let $\phi(\psi(\cdot, u), r(u)) : \ell^\infty(\Theta)^p \times \mathbb{R} \mapsto \Theta$ be a deterministic map that assigns one of its $r(u)$-approximate zeroes to each element $\psi(\cdot, u) \in \ell^\infty(\Theta)^p$. Further, in our case $\psi(\cdot, u)$'s are all indexed by u, and so we can think of $\psi = (\psi(\theta, u) : u \in \mathcal{U})$ as an element of $\ell^\infty(\Theta \times \mathcal{U})^p$, and of $r = (r(u) : u \in \mathcal{U})$ as an element of $\ell^\infty(\mathcal{U})$. Then we can define $\phi(\psi, r)$ as a map that assigns a function $u \mapsto \phi(\psi(\cdot, u), r(u))$ to each element (ψ, r). The properties of the Z-processes therefore rely on Hadamard differentiability of the Z-map

$$(\psi, r) \mapsto \phi(\psi, r)$$

at $(\psi, r) = (\Psi, 0)$, i.e. differentiability with respect to the underlying vector of moments function and with respect to numerical tolerance parameter r.

We make the following assumption about the vector of moment functions. Let $B_\delta(\theta)$ denote an open ball of radius δ centered at θ.

CONDITION Z. *Let $\mathcal{U}$ be a compact set of some metric space, and Θ be an arbitrary subset of $\mathbb{R}^p$. Assume (i) for each $u \in \mathcal{U}$, $\Psi(\cdot, u) : \Theta \mapsto \mathbb{R}^p$ possesses a unique zero at $\theta_0(u)$,*

36

and $\mathcal{N} = \cup_{u \in \mathcal{U}} B_\delta(\theta_0(u))$ is a relatively compact subset of Θ for some $\delta > 0$, (ii) The inverse of $\Psi(\cdot, u)$ defined as $\Psi^{-1}(x, u) := \{\theta \in \Theta : \Psi(\theta, u) = x\}$ is continuous at $x = 0$ uniformly in $u \in \mathcal{U}$ with respect to Hausdorff distance, (iii) there exists $\dot{\Psi}_{\theta_0(u),u}$ such that $\lim_{t \searrow 0} \sup_{u \in \mathcal{U}, \|h\|=1} |t^{-1}(\Psi(\theta_0(u) + th, u) - \Psi(\theta_0(u), u)) - \dot{\Psi}_{\theta_0(u),u} h| = 0$, where $\inf_{u \in \mathcal{U}} \inf_{\|h\|=1} \|\dot{\Psi}_{\theta_0(u),u} h\| > 0$, and (iv) the maps $u \mapsto \theta_0(u)$ and $u \mapsto \dot{\Psi}_{\theta_0(u),u}$ are continuous.

The following lemma is useful for verifying Condition Z.

Lemma E.1 (Simple sufficient condition for Z). *Suppose that $\Theta = \mathbb{R}^p$, and $\mathcal{U}$ is a compact interval in $\mathbb{R}$. Let $\mathcal{I}$ be an open set containing $\mathcal{U}$. (a) $\Psi : \Theta \times \mathcal{I} \mapsto \mathbb{R}^p$ is continuous, and $\theta \mapsto \Psi(\theta, u)$ is the gradient of a convex function in θ for each $u \in \mathcal{U}$, (b) for each $u \in \mathcal{U}$, $\Psi(\theta_0(u), u) = 0$, (c) $\frac{\partial}{\partial(\theta', u)} \Psi(\theta, u)$ exists at $(\theta_0(u), u)$ and is continuous at $(\theta_0(u), u)$ for each $u \in \mathcal{U}$, and $\dot{\Psi}_{\theta_0(u),u} := \frac{\partial}{\partial \theta'} \Psi(\theta, u)|_{\theta_0(u)}$ obeys $\inf_{u \in \mathcal{U}} \inf_{\|h\|=1} \|\dot{\Psi}_{\theta_0(u),u} h\| > c_0 > 0$. Then Condition Z holds and $u \mapsto \theta_0(u)$ is continuously differentiable.*

Lemma E.2 (Hadamard differentiability of approximate Z-maps). *Suppose that Condition Z(i)-(iii) holds. Then, the map $(\psi, r) \mapsto \phi(\psi, r)$ is Hadamard differentiable at $(\psi, r) = (\Psi, 0)$ as a map $\phi : \mathbb{D} = \ell^\infty(\Theta \times \mathcal{U})^p \times \ell^\infty(\mathcal{U}) \mapsto \mathbb{E} = \ell^\infty(\mathcal{U})^p$ tangentially to $\mathbb{D}_0 = \mathbb{D} \cap (C(\mathcal{N} \times \mathcal{U})^p \times \{0\})$, with the derivative map $(z, 0) \mapsto \phi'_{\Psi,0}(z, 0)$ defined by*

$$\phi'_{\Psi,0}(z, 0) = -\dot{\Psi}^{-1}_{\theta_0(\cdot),\cdot} z(\theta_0(\cdot), \cdot),$$

where $(z, 0) \mapsto \phi'_{\Psi,0}(z, 0)$ is defined and continuous over $z \in \ell^\infty(\Theta \times \mathcal{U})^p$. If in addition Condition Z(iv) holds, then $u \mapsto -\dot{\Psi}^{-1}_{\theta_0(u),u} z(\theta_0(u), u)$ is continuous.

This lemma is an alternative to Lemma 3.9.34 in VW on Hadamard differentiability of Z-maps in general normed spaces, which we found difficult to use in our case. (The paths of quantile regression processes $\widehat{\theta}(\cdot)$ in the non-univariate case are somewhat irregular and it is not apparent how to place them in an entropically simple parameter space.) Moreover, our lemma applies to approximate Z-estimators. This allows us to cover quantile regression processes, where exact Z-estimators do not exist for *any* sample size. The following lemma shows how to apply Lemma E.2 to a generic Z-problem.

Lemma E.3 (Limit distribution for approximate Z-estimators). *Suppose condition Z(i)-(iii) holds. If $\sqrt{n}(\widehat{\Psi} - \Psi) \rightsquigarrow Z$ in $\ell^\infty(\Theta \times \mathcal{T})^p$, where Z is a Gaussian process with a.s. continuous paths on $\mathcal{N} \times \mathcal{U}$, and $\|n^{1/2} \widehat{r}\|_{\mathcal{U}} \to_{\mathbb{P}} 0$, then*

$$\sqrt{n}(\widehat{\theta}(\cdot) - \theta_0(\cdot)) = -\dot{\Psi}^{-1}_{\theta_0(\cdot),\cdot} \sqrt{n}(\widehat{\Psi} - \Psi)(\theta_0(\cdot), \cdot) + o_{\mathbb{P}}(1) \rightsquigarrow -\dot{\Psi}^{-1}_{\theta_0(\cdot),\cdot} [Z(\theta_0(\cdot), \cdot)] \text{ in } \ell^\infty(\mathcal{U})^p.$$

If Condition Z(iv) also holds, then the paths $u \mapsto -\dot{\Psi}^{-1}_{\theta_0(u),u} [Z(\theta_0(u), u)]$ are continuous, a.s. Moreover, if $\sqrt{n}(\widehat{\Psi}^* - \widehat{\Psi}) \rightsquigarrow_{\mathbb{P}} Z$ in $\ell^\infty(\Theta \times \mathcal{U})^p$, and $\|n^{1/2} \widehat{r}^*\|_{\mathcal{U}} \to_{\mathbb{P}} 0$, then

$$\sqrt{n}(\widehat{\theta}^*(\cdot) - \widehat{\theta}(\cdot)) \rightsquigarrow_{\mathbb{P}} -\dot{\Psi}^{-1}_{\theta_0(\cdot),\cdot} [Z(\theta_0(\cdot), \cdot)] \text{ in } \ell^\infty(\mathcal{U})^p.$$

E.2. **Proofs of Lemma E.1-E.3. Proof of Lemma E.1.** To show Condition Z(i), note that for each $u \in \mathcal{U}$, $\Psi(\cdot, u) : \Theta \mapsto \mathbb{R}^p$ possesses a unique zero at $\theta_0(u)$ by conditions (a) and (b). By the Implicit Function Theorem, $\partial \theta_0(u)/\partial u = -\dot{\Psi}_{\theta_0(u),u}^{-1} \times [\partial \Psi(\theta_0(u), u)/\partial u]$, which is uniformly bounded and continuous in $u \in \mathcal{U}$ by condition (c) and compactness of $\mathcal{U}$. Hence $\mathcal{N} = \cup_{u \in \mathcal{U}} B_\delta(\theta_0(u))$ is a compact subset of Θ for any $\delta > 0$. This verifies Condition Z(i) and also implied condition Z(iv) in view of condition (c) and continuous differentiability (and hence continuity) of $u \mapsto \theta_0(u)$.

To show Condition Z(iii), take any sequence $(u_t, h_t) \to (u, h)$ with $u \in \mathcal{U}, h \in \mathbb{R}^p$ and then note that, for $t^* \in [0, t]$, $\Delta_t(u_t, h_t) = t^{-1}\{\Psi(\theta(u_t) + th_t, u_t) - \Psi(\theta(u_t), u_t)\} = \frac{\partial \Psi}{\partial \theta}(\theta(u_t) + t^* h_t, u_t)h_t \to \frac{\partial \Psi}{\partial \theta}(\theta_0(u), u)h = \dot{\Psi}_{\theta_0(u),u}h$ using the continuity characterizations of the derivative $\partial \Psi/\partial \theta$ and the continuity of $u \mapsto \theta_0(u)$ established in the first paragraph. Hence by Lemma B.4, we conclude that $\sup_{u \in \mathcal{U}, \|h\|=1} |\Delta_t(u, h) - \dot{\Psi}_{\theta_0(u),u}h| \to 0$ as $t \searrow 0$.

To show Condition Z(ii), we need to verify that for any $x_t \to 0$ such that $x_t \in \Psi(\Theta, u)$, $d_H(\Psi^{-1}(x_t, u), \Psi^{-1}(0, u)) \to 0$, where d_H is the Hausdorff distance, uniformly in $u \in \mathcal{U}$. Suppose by contradiction that this is not true, then there is (x_t, u_t) with $x_t \to 0$ and $u_t \in \mathcal{U}$ such that $d_H(\Psi^{-1}(x_t, u_t), \Psi^{-1}(0, u_t)) \not\to 0$. By compactness of $\mathcal{U}$, we can select a further subsequence (x_k, u_k) such that $u_k \to u$, where $u \in \mathcal{U}$. We have that $\Psi^{-1}(0, u) = \theta_0(u)$ is continuous in $u \in \mathcal{U}$, so we must have $d_H(\Psi^{-1}(x_k, u_k), \Psi^{-1}(0, u)) \not\to 0$. Hence there is a further subsequence $y_l \in \Psi^{-1}(x_l, u_l)$ with $y_l \to y$ in $\overline{\mathbb{R}}^p$, such that $y \neq \Psi^{-1}(0, u) = \theta_0(u)$, and such that $x_l = \Psi(y_l, x_l) \to 0$. If $y \in \mathbb{R}^p$, by continuity $\Psi(y_l, u_l) \to \Psi(y, u) \neq 0$ since $y \neq \Psi^{-1}(0, u)$, yielding a contradiction. If $y \in \overline{\mathbb{R}}^p \backslash \mathbb{R}^p$, we need to show that $\|\Psi(y_l, u_l)\| \not\to 0$ to obtain a contradiction. Note that for $h \in \mathbb{R}^p$: $\|h\| = 1$, $u \in \mathcal{U}$, and scalar $\delta \in \mathbb{R}$, the map $\delta \mapsto \Psi(\theta_0(u) + \delta h, u)'h$ is non-decreasing by $\theta \mapsto \Psi(\theta)$ being the gradient of a convex function. Since $\Psi(\theta_0(u), u) = 0$, conclude that $|\Psi(\theta_0(u) + \delta h, u)'h|$ is non-decreasing in $|\delta|$. Moreover, $\|\Psi(\theta_0(u) + \delta h, u)\| \geq |\Psi(\theta_0(u) + \delta h, u)'h|$ for any (h, u, δ). Hence to get contradiction it suffices to show that $\inf_{u \in \mathcal{U}, \|h\|=1} |\Psi(\theta_0(u) + \delta h, u)'h| > 0$ for some $\delta > 0$. Indeed, for small enough $\delta > 0$, by computation similar to that above and condition (c), this quantity is bounded below by $(1/2)\delta \inf_{u \in \mathcal{U}} \inf_{\|h\|=1} |h'\dot{\Psi}_{\theta_0(u),u}h| \geq c_0\delta/2 > 0$. $\qquad \square$

Proof of Lemma E.2. Consider $\psi_t = \Psi + tz_t$ and $r_t = 0 + tq_t$ with $z_t \to z$ in $\ell^\infty(\Theta \times \mathcal{U})^p$ where $z \in C(\mathcal{N} \times \mathcal{U})^p$ and $q_t \to 0$ in $\ell^\infty(\mathcal{U})$. Then, for $\theta_t(\cdot) = \phi(\psi_t, r_t)$ we need to prove that uniformly in $u \in \mathcal{U}$,

$$\frac{\theta_t(u) - \theta_0(u)}{t} \to \phi'_{\Psi,0}(z, 0)(u) = -\dot{\Psi}_{\theta_0(u),u}^{-1}[z(\theta_0(u), u)].$$

We have that $\Psi(\theta_0(u), u) = 0$ for all $u \in \mathcal{U}$. By definition, $\theta_t(u)$ satisfies

$$\|\Psi(\theta_t(u), u) - \Psi(\theta_0(u), u) + tz_t(\theta_t(u), u)\|^2 \leq \inf_{\theta \in \Theta} \|\Psi(\theta, u) + tz_t(\theta, u)\|^2 + t^2 q_t^2(u) =: t^2 \lambda_t^2(u) + t^2 q_t^2(u),$$

uniformly in $u \in \mathcal{U}$. The rest of the proof has three steps. In Step 1, we establish a rate of convergence of $\theta_t(\cdot)$ to $\theta_0(\cdot)$. In Step 2, we verify the main claim of the lemma concerning the linear representation for $t^{-1}(\theta_t(\cdot) - \theta_0(\cdot))$, assuming that $\lambda_t(\cdot) = o(1)$. In Step 3, we verify that $\lambda_t(\cdot) = o(1)$.

38

STEP 1. Here we show that uniformly in $u \in \mathcal{U}$, $\|\theta_t(u) - \theta_0(u)\| = O(t)$. First observe that $\sup_{(\theta,u) \in \Theta \times \mathcal{U}} \|z_t(\theta, u)\| = O(1)$ by $z_t \to z$ and $\sup_{(\theta,u) \in \Theta \times \mathcal{U}} \|z(\theta, u)\| < \infty$. Then note that $\lambda_t(u) \le \|t^{-1}\Psi(\theta_0(u), u) + z_t(\theta_0(u), u)\| = \|z(\theta_0(u), u) + o(1)\| = O(1)$ uniformly in $u \in \mathcal{U}$. We conclude that uniformly in $u \in \mathcal{U}$, as $t \searrow 0$: $t^{-1}(\Psi(\theta_t(u), u) - \Psi(\theta_0(u), u)) = -z_t(\theta_t(u), u) + O(\lambda_t(u) + q_t(u)) = O(1)$ and $\|\Psi(\theta_t(u), u) - \Psi(\theta_0(u), u)\| = O(t)$. By assumption $\Psi(\cdot, u)$ has a unique zero at $\theta_0(u)$ and has an inverse that is continuous at zero uniformly in $u \in \mathcal{U}$; hence it follows that uniformly in $u \in \mathcal{U}$, $\|\theta_t(u) - \theta_0(u)\| \le d_H(\Psi^{-1}(\Psi(\theta_t(u), u), u), \Psi^{-1}(0, u)) \to 0$, where d_H is the Hausdorff distance. By condition Z(iii) uniformly in $u \in \mathcal{U}$

$$\liminf_{t \searrow 0} \frac{\|\Psi(\theta_t(u), u) - \Psi(\theta_0(u), u)\|}{\|\theta_t(u) - \theta_0(u)\|} \ge \liminf_{t \searrow 0} \frac{\|\dot{\Psi}_{\theta_0(u),u}[\theta_t(u) - \theta_0(u)]\|}{\|\theta_t(u) - \theta_0(u)\|}$$
$$\ge \inf_{\|h\|=1} \|\dot{\Psi}_{\theta_0(u),u} h\| = c > 0,$$

where h ranges over $\mathbb{R}^p$, and $c > 0$ by assumption. The claim of the step follows.

STEP 2. (Main) Here we verify the main claim of the lemma. Using Condition Z(iii) again, conclude $\|\Psi(\theta_t(u), u) - \Psi(\theta_0(u), u) - \dot{\Psi}_{\theta_0(u),u}[\theta_t(u) - \theta_0(u)]\| = o(t)$ uniformly in $u \in \mathcal{U}$. Below we show that $\lambda_t(u) = o(1)$ and we also have $q_t(u) = o(1)$ uniformly in $u \in \mathcal{U}$ by assumption. Thus, we can conclude that uniformly in $u \in \mathcal{U}$, $t^{-1}(\Psi(\theta_t(u), u) - \Psi(\theta_0(u), u)) = -z_t(\theta_t(u), u) + o(1) = -z(\theta_0(u), u) + o(1)$ and

$$t^{-1}[\theta_t(u) - \theta_0(u)] = \dot{\Psi}_{\theta_0(u),u}^{-1} \left[t^{-1}(\Psi(\theta_t(u), u) - \Psi(\theta_0(u), u)) + o(1)\right]$$
$$= -\dot{\Psi}_{\theta_0(u),u}^{-1}[z(\theta_0(u), u)] + o(1).$$

STEP 3. In this step we show that $\lambda_t(u) = o(1)$ uniformly in $u \in \mathcal{U}$. Note that for $\bar{\theta}_t(u) := \theta_0(u) - t\dot{\Psi}_{\theta_0(u),u}^{-1}[z(\theta_0(u), u)] = \theta_0(u) + O(t)$, we have that $\bar{\theta}_t(u) \in \mathcal{N} = \cup_{u \in \mathcal{U}} B_\delta(\theta_0(u))$, for small enough t, uniformly in $u \in \mathcal{U}$; moreover, $\lambda_t(u) \le \|t^{-1}\Psi(\bar{\theta}_t(u), u) + z_t(\bar{\theta}_t(u), u)\|$ which is equal to $\| - \dot{\Psi}_{\theta_0(u),u}\{\dot{\Psi}_{\theta_0(u),u}^{-1}[z(\theta_0(u), u)]\} + z(\theta_0(u), u) + o(1)\| = o(1)$, as $t \searrow 0$. $\qquad \square$

Proof of Lemma E.3. We shall omit the dependence on u signified by $(\cdot)$ in what follows. Then, in the notation of Lemma E.2, $\widehat{\theta} = \phi(\widehat{\Psi}, \widehat{r})$ is an estimator of $\theta_0 = \phi(\Psi, 0)$. By the Hadamard differentiability of the ϕ-map shown in Lemma E.2, the weak convergence conclusion follows. The first order expansion follows by noting that the linear map $\psi \mapsto -\dot{\Psi}_{\theta_0}^{-1}\psi$ is trivially Hadamard differentiable at $\psi = \Psi$, and so by stacking, $(-\sqrt{n}(\widehat{\theta} - \theta_0), \dot{\Psi}_{\theta_0}^{-1}\sqrt{n}(\widehat{\Psi} - \Psi)) \rightsquigarrow (\dot{\Psi}_{\theta_0}^{-1}Z, \dot{\Psi}_{\theta_0}^{-1}Z)$ in $\ell^\infty(\mathcal{U})^{2p}$, and so the difference between the terms converges in outer probability to zero. The validity of bootstrap follows from the delta method for the bootstrap. $\qquad \square$

E.3. **Limits of empirical measures.** The following result is useful to organize thoughts for the case of transformation sampling. Let

$$\widehat{\mathbb{G}}_k(f) := \frac{1}{\sqrt{n_k}} \sum_{i=1}^{n} \left(f(Y_{ki}, X_{ki}) - \int f dP_k\right) \quad \text{and} \quad \widehat{\mathbb{G}}_k^*(f) := \frac{1}{\sqrt{n_k}} \sum_{i=1}^{n} (w_{ki} - \bar{w}_k) f(Y_{ki}, X_{ki})$$

be the empirical and exchangeable bootstrap processes for the sample from population k.

Lemma E.4. *Suppose Conditions S, SM, and EB hold. Let $\mathcal{F} : \mathcal{X} \mapsto \mathbb{R}$ be a Dudley-Pollard class (as defined in notation section), where $\mathcal{X} \supseteq \cup_{k \in \mathcal{K}} \mathcal{X}_k$. (1) Then $\widehat{\mathbb{G}}_k(f) \rightsquigarrow \mathbb{G}_k(f)$ and $\widehat{\mathbb{G}}_k^*(f) \rightsquigarrow_{\mathbb{P}} \mathbb{G}_k(f)$ as stochastic processes indexed by $(k, f) \in \mathcal{K}_0 \mathcal{F}$ in $\ell^\infty(\mathcal{K}_0 \mathcal{F})$. (2) Moreover, $\widehat{\mathbb{G}}_k(f) \rightsquigarrow \mathbb{G}_k(f)$ and $\widehat{\mathbb{G}}_k^*(f) \rightsquigarrow_{\mathbb{P}} \mathbb{G}_k(f)$ as stochastic processes indexed by $(k, f) \in \mathcal{K} \mathcal{F}$ in $\ell^\infty(\mathcal{K} \mathcal{F})$, where $\mathbb{G}_k(f) = \mathbb{G}_{(k)}(f \circ g_{(k),k})$, provided that $\mathcal{F} \circ g_{l(k),k}$ continues to be a Dudley-Pollard class for all $k \in \mathcal{K}_t$.*

Proof of Lemma E.4. Note that $\mathcal{F}$ is a universal Donsker class by Dudley (1987). Statement (1) then follows from the independence of samples across $k \in \mathcal{K}_0$, so that joint convergence follows from the marginal convergence for each $k \in \mathcal{K}_0$, and from the results on exchangeable bootstrap given in Chapter 3.6 of VW. Let $\mathcal{F}$ be a Dudley-Pollard class. To show Statement (2) we note that $\widehat{\mathbb{G}}_k(f) = \widehat{\mathbb{G}}_m(f \circ g_{m,k})$ for $m = l(k) \in \mathcal{K}_0$. Recall that $l(\cdot)$ denotes the indexing function that indicates the population $l(k)$ from which the k-th population is created by transformation, in particular $l(m) = m$ if $m \in \mathcal{K}_0$. Thus, $l^{-1}(m) = \{k \in \mathcal{K} : l(k) = m\}$ is the set of all populations created from the m-th population that includes m itself. Let $\mathcal{F}'$ consist of $\mathcal{F}$ and $\mathcal{F} \circ g_{m,k}$ for all $k \in l^{-1}(m) = \{m, ...\} \subset \mathcal{K}$ and all $m \in \mathcal{K}_0$. Then $\mathcal{F}'$ is a Dudley-Pollard class, since it is a finite union of Dudley-Pollard classes ($\mathcal{K}$ is finite), so statement (2) follows from statement (1). In fact, this shows that the convergence analysis is reducible to the independent case by suitably enriching $\mathcal{F}$ into the class $\mathcal{F}'$. $\qquad\square$

E.4. **Proof of Theorem 5.1.** (Validity of QR based Counterfactual Analysis) The proof of preceding lemma shows that by suitably enlarging the class $\mathcal{F}$, it suffices to consider only the independent samples, i.e. those with population indices $k \in \mathcal{K}_0$. Moreover, by independence across k, the joint convergence result follows from the marginal convergence for each k separately. It remains to examine each case with $k \in \mathcal{J}$ separately, since otherwise for a given $k \notin \mathcal{J}$, the convergence of empirical measures and associated bootstrap result are already shown in Lemma E.4. In what follows, since the proof can be done for each k marginally, we shall omit the index k to simplify the notation.

STEP 1.(Results for coefficients and empirical measures). Let $\mathcal{F}$ be a Dudley-Pollard class, as defined in notation section. We use the Z-process framework described above, where we let $\theta(u) = \beta(u)$, and $\Theta = \mathbb{R}^{d_x}$. Lemma E.3 above illustrates the use of the delta method for a single Z-estimation problem, which the reader may find helpful before reading this proof. Let $\varphi_{u,\beta}(Y, X) = (1\{Y \leq X'\beta\} - u)X$, $\Psi(\theta, u) = P[\varphi_{u,\beta}]$, and $\widehat{\Psi}(\theta, u) = P_n[\varphi_{u,\beta}]$, where P_n is the empirical measure and P is the corresponding probability measure. From the subgradient characterization, we know that the QR estimator obeys $\widehat{\beta}(u) = \phi(\widehat{\Psi}(\cdot, u), \widehat{r}(u))$, $\widehat{r}(u) = \max_{1 \leq i \leq n} \|X_i\| d_x / n$, for each $u \in \mathcal{U}$, with $n^{1/2} \|\widehat{r}\|_{\mathcal{U}} \rightarrow_{\mathbb{P}} 0$, where ϕ is an approximate Z- map as defined in Appendix E.1. The random vector $\widehat{\beta}(u)$ and $\int f d\widehat{F}_X = P_n(f)$ are estimators of $\beta(u) = \phi(\Psi(\cdot, u), 0)$ and $\int f dF_X = P(f)$. Then, by Step 3 below

$$(\sqrt{n}(\widehat{\Psi} - \Psi), \widehat{\mathbb{G}}) \rightsquigarrow (W, \mathbb{G}) \text{ in } \ell^\infty(\mathbb{R}^{d_x} \times \mathcal{U})^{d_x} \times \ell^\infty(\mathcal{F}), \ W(\beta, u) = \mathbb{G}\varphi_{u,\beta},$$

where W has continuous paths a.s. Step 4 verifies the conditions of Lemma E.1 for $\dot{\Psi}_{\theta_0(u),u} = J(u)$, thereby also implying continuous differentiability of $u \mapsto \beta(u)$. Then, by Lemma E.2, the map ϕ is Hadamard-differentiable with derivative map $w \mapsto -J^{-1}w$ at $(\Psi, 0)$. Therefore, we can conclude by the Functional Delta Method that $(\sqrt{n}(\widehat{\beta}(\cdot) - \beta(\cdot)), \widehat{\mathbb{G}}) \rightsquigarrow (-J^{-1}(\cdot)W(\beta(\cdot), \cdot), \mathbb{G})$ in $\ell^\infty(\mathcal{U})^{d_x} \times \ell^\infty(\mathcal{F})$, where $-J^{-1}(\cdot)W(\beta(\cdot), \cdot)$ has continuous paths a.s.

Similarly, for the bootstrap version, we have from the subgradient characterization of the QR estimator that $\widehat{\beta}^*(u) = \phi(\widehat{\Psi}(\cdot, u), \widehat{r}^*(u))$, $\widehat{r}^*(u) = \max_i w_i \|X_i\| d_x / n$, with $n^{1/2}\widehat{r}_n^* \to_{\mathbb{P}} 0$ and hence also $\rightsquigarrow_{\mathbb{P}} 0$, by $E[\max_{i \leq n} w_i \|X_i\| / \sqrt{n}] = o(1)$, which holds since $E\|w_i X_i\|^{2+\epsilon} = E|w_i|^{2+\epsilon} E\|X_i\|^{2+\epsilon} < \infty$ by assumption on the bootstrap weights and $\|X_i\|$. By Step 3 below, $(\sqrt{n}(\widehat{\Psi}^* - \widehat{\Psi}), \widehat{\mathbb{G}}^*) \rightsquigarrow_{\mathbb{P}} (W, \mathbb{G})$ in $\ell^\infty(\mathbb{R}^{d_x} \times \mathcal{U})^{d_x} \times \ell^\infty(\mathcal{F})$. Therefore by the Functional Delta method for Bootstrap $(\sqrt{n}(\widehat{\beta}^*(\cdot) - \widehat{\beta}(\cdot)), \widehat{\mathbb{G}}^*) \rightsquigarrow_{\mathbb{P}} (-J^{-1}(\cdot)W(\beta(\cdot), \cdot), \mathbb{G})$ in $\ell^\infty(\mathcal{U})^{d_x} \times \ell^\infty(\mathcal{F})$. Hence the conclusion (2) stated in Corollary 5.1 follows.

STEP 2.(Main: Results for conditional cdfs). Here we shall rely on compactness of $\mathcal{YX}$. In order to verify Condition D, we first note that $\mathcal{F}_0 = \{F_{Y|X}(y|\cdot) : y \in \mathcal{Y}\}$ is a uniformly bounded "parametric" family indexed by $y \in \mathcal{Y}$ that obeys $|F_{Y|X}(y|\cdot) - F_{Y|X}(y'|\cdot)| \leq L|y - y'|$, given the assumption that the density function $f_{Y|X}$ is uniformly bounded by some constant L. Given compactness of $\mathcal{Y}$, the uniform ϵ-covering numbers for this class can be bounded independently of F_X by const/ϵ, and so Pollard's entropy integral. Hence we can construct a class of functions $\mathcal{F}$ containing the union of all the families $\mathcal{F}_0$ for the populations in $\mathcal{J}$ and the indicators of all the rectangles in $\mathbb{R}^{d_x}$. Note that these indicators form a VC class. The final set $\mathcal{F}$ therefore is a Dudley-Pollard class.

Next consider the mapping $\varphi : \mathbb{D}_\varphi \subset \ell^\infty(\mathcal{U})^{d_x} \mapsto \ell^\infty(\mathcal{YX})$, defined as $b \mapsto \varphi(b)$, $\varphi(b)(y, x) = \varepsilon + \int_\varepsilon^{1-\varepsilon} 1\{x'b(u) \leq y\}du$. It follows from the results of Chernozhukov, Fernandez-Val, and Galichon (2010) that this map is Hadamard differentiable at $b(\cdot) = \beta(\cdot)$ tangentially to $C(\mathcal{U})^{d_x}$, with the derivative map given by: $\alpha \mapsto \varphi'_{\beta(\cdot)}(\alpha)$, $\varphi'_{\beta(\cdot)}(\alpha)(y, x) = f_{Y|X}(y|x)x'\alpha(F_{Y|X}(y|x))$. Since $\widehat{F}_{Y|X} = \varphi(\widehat{\beta}(\cdot))$ and $\int f d\widehat{F}_X = \int f dP_n$ are estimators of $F_{Y|X} = \varphi(\beta(\cdot))$ and $\int f dF_X = \int f dP$, by the delta method it follows that

$$(\sqrt{n}(\widehat{F}_{Y|X} - F_{Y|X}), \widehat{\mathbb{G}}) \rightsquigarrow (-\varphi'_{\beta(\cdot)}J^{-1}(\cdot)W(\cdot, \beta(\cdot)), \mathbb{G}) \text{ in } \ell^\infty(\mathcal{YX}) \times \ell^\infty(\mathcal{F}), \quad \text{(E.1)}$$

$$(\sqrt{n}(\widehat{F}^*_{Y|X} - \widehat{F}_{Y|X}), \widehat{\mathbb{G}}^*) \rightsquigarrow_{\mathbb{P}} (-\varphi'_{\beta(\cdot)}J^{-1}(\cdot)W(\cdot, \beta(\cdot)), \mathbb{G}) \text{ in } \ell^\infty(\mathcal{YX}) \times \ell^\infty(\mathcal{F}). \quad \text{(E.2)}$$

STEP 3. (Auxiliary: Donskerness). First, we note that $\{\varphi_{u,\beta}(Y, X) : (u, \beta) \in \mathcal{U} \times \mathbb{R}^{d_x}\}$ is P-Donsker. This follows by a standard argument, which is omitted. Second, we note that $(u, \beta) \mapsto \varphi_{u,\beta}(Y, X)$ is $L^2(P)$ continuous by the dominated convergence theorem, and the fact that $(\beta, u) \mapsto (1(Y \leq X'\beta) - u)X$ is continuous at each $(\beta, u) \in \mathbb{R}^{d_x} \times \mathcal{U}$ with probability one by the absolute continuity of $F_{Y|X}$, and its norm is bounded by a square integrable function $2\|X\|$ under P. Hence $\mathbb{G}(\varphi_{u,\beta})$ has continuous paths in (u, β) and the convergence results follow from the convergence results in Lemma E.4.

STEP 4. (Auxiliary: Verification of Conditions of Lemma E.1).We verify conditions (a)-(c) of Lemma E.1. Conditions (a) and (b) are immediate by the assumptions. To verify (c), we can compute $\frac{\partial}{\partial(b',u)}\Psi(b,u) = [E[f_{Y|X}(X'b|X)XX'], -EX]$ for (b,u) in the neighborhood of $(\beta(u), u)$, where the right side is continuous at $(b,u) = (\beta(u), u)$ for each $u \in \mathcal{U}$. This computation and continuity follows from using the dominated convergence theorem, the a.s. continuity and uniform boundedness of the mapping $y \mapsto f_{Y|X}(y|X)$, as well as $E\|X\|^2 < \infty$. By assumption, the minimum eigenvalue of $J(u) = E[f_{Y|X}(X'\beta(u)|X)XX']$ is bounded away from zero uniformly in $u \in \mathcal{U}$. $\qquad\square$

E.5. **Proof of Theorem 5.2.** (Validity of DR based Counterfactual Analysis). As in the previous proof, it suffices to show the result for each $k \in \mathcal{J}$ separately. In what follows, since the proof can be done for each k marginally, we shall omit the index k to simplify the notation. We only consider the case where $\mathcal{Y}$ is a compact interval of $\mathbb{R}$. The case where $\mathcal{Y}$ is finite is simpler and follows similarly.

STEP 1.(Results for coefficients and empirical measures). We use the Z-process framework described above, where we let $u = y, \theta(u) = \beta(y)$, $\Theta = \mathbb{R}^{d_x}$, and $\mathcal{U} = \mathcal{Y}$. Lemma E.3 above illustrates the use of the delta method for a single Z-estimation problem, which the reader may find helpful before reading this proof. Let

$$\varphi_{y,\beta}(Y, X) = [\Lambda(X'\beta) - 1(Y \leq y)]H(X'\beta)X,$$

where $H(z) = \lambda(z)/\{\Lambda(z)[1 - \Lambda(z)]\}$ and λ is the derivative of Λ. Let $\Psi(\theta, y) = P[\varphi_{y,\beta}]$ and $\widehat{\Psi}(\theta, u) = P_n[\varphi_{y,\beta}]$, where P_n is the empirical measure and P is the corresponding probability measure. From the first order conditions, distribution regression in the sample obeys $\widehat{\beta}(y) = \phi(\widehat{\Psi}(\cdot, u), 0)$, for each $y \in \mathcal{Y}$, where ϕ is the Z- map defined in Appendix E.1. The random vector $\widehat{\beta}(y)$ and $\int f d\widehat{F}_X = P_n(f)$ are estimators of $\beta(y) = \phi(\Psi(\cdot, u), 0)$ and $\int f dF_X = P(f)$. Then, by Step 3 below

$$(\sqrt{n}(\widehat{\Psi} - \Psi), \widehat{\mathbb{G}}) \rightsquigarrow (W, \mathbb{G}) \text{ in } \ell^\infty(\mathbb{R}^{d_x} \times \mathcal{Y})^{d_x} \times \ell^\infty(\mathcal{F}), \ W(y, \beta) = \mathbb{G}\varphi_{y,\beta},$$

where W has continuous paths a.s. Step 4 verifies the Conditions of Lemma E.1 for $\dot{\Psi}_{\theta_0(u),u} = J(y)$, which also implies that $y \mapsto \beta(y)$ is continuously differentiable on the interval $\mathcal{Y}$. Then, by Lemma E.2, the map ϕ is Hadamard-differentiable with the derivative map $w \mapsto -J^{-1}w$ at $(\Psi, 0)$. Therefore, we can conclude by the Functional Delta Method that

$$(\sqrt{n}(\widehat{\beta}(\cdot) - \beta(\cdot)), \widehat{\mathbb{G}}) \rightsquigarrow (-J^{-1}(\cdot)W(\beta(\cdot), \cdot), \mathbb{G}) \text{ in } \ell^\infty(\mathcal{Y})^{d_x} \times \ell^\infty(\mathcal{F}),$$

where $-J^{-1}(\cdot)W(\beta(\cdot), \cdot)$ has continuous paths a.s.

Similarly, for the bootstrap version, we have from the first order conditions of the DR estimator that $\widehat{\beta}^*(y) = \phi(\widehat{\Psi}(\cdot, u), 0)$, and $(\sqrt{n}(\widehat{\Psi}^* - \widehat{\Psi}), \widehat{\mathbb{G}}^*) \rightsquigarrow_\mathbb{P} (W, \mathbb{G})$ in $\ell^\infty(\mathbb{R}^{d_x} \times \mathcal{Y})^{d_x} \times \ell^\infty(\mathcal{F})$ by Step 3 below. Therefore by the Functional Delta method for Bootstrap

$$(\sqrt{n}(\widehat{\beta}^*(\cdot) - \widehat{\beta}(\cdot)), \widehat{\mathbb{G}}^*) \rightsquigarrow_\mathbb{P} (-J^{-1}(\cdot)W(\cdot, \beta(\cdot)), \mathbb{G}) \text{ in } \ell^\infty(\mathcal{Y})^{d_x} \times \ell^\infty(\mathcal{F}).$$

Hence the conclusion (2) stated in Corollary 5.1 follows. The first-order expansion of the conclusion (1) in Corollary 5.1 follows by an argument similar to the proof of Lemma E.3.

STEP 2.(Main: Results for conditional cdfs). Here we shall rely on compactness of $\mathcal{Y}\mathcal{X}$. Then, $\mathcal{Y}$ is a closed interval of $\mathbb{R}$. In order to verify Condition D, we first note that $\mathcal{F}_0 = \{F_{Y|X}(y|\cdot) : y \in \mathcal{Y}\}$ is a uniformly bounded "parametric" family indexed by $y \in \mathcal{Y}$ that obeys $|F_{Y|X}(y|\cdot) - F_{Y|X}(y'|\cdot)| \leq L|y - y'|$, given the assumption that the density function $f_{Y|X}$ is uniformly bounded by some constant L. Given compactness of $\mathcal{Y}$, the uniform ϵ-covering numbers for this class can be bounded independently of F_X by const/ϵ, and so the Pollard's entropy integral is finite. Hence we can construct a class of functions $\mathcal{F}$ containing the union of all the families $\mathcal{F}_0$ for the populations in $\mathcal{J}$ and the indicators of all rectangles in $\overline{\mathbb{R}}^{d_x}$. Note that these indicators form a VC class. The final set $\mathcal{F}$ thereforeis a Dudley-Pollard class.

Next consider the mapping $\varphi : \mathbb{D}_\varphi \subset \ell^\infty(\mathcal{Y})^{d_x} \mapsto \ell^\infty(\mathcal{Y}\mathcal{X})$, defined as $b \mapsto \varphi(b)$, $\varphi(b)(x,y) = \Lambda(x'b(y))$. It is straightforward to deduce that this map is Hadamard differentiable at $b(\cdot) = \beta(\cdot)$ tangentially to $C(\mathcal{Y})^{d_x}$ with the derivative map given by: $\alpha \mapsto \varphi'_{\beta(\cdot)}(\alpha)$, $\varphi'_{\beta(\cdot)}(\alpha)(y,x) = \lambda(x'\beta(y))x'\alpha(y)$. Since $\widehat{F}_{Y|X} = \varphi(\widehat{\beta}(\cdot))$ and $\int f d\widehat{F}_X = \int f dP_n$ are estimators of $F_{Y|X} = \varphi(\beta(\cdot))$ and $\int f dF_X = \int f dP$, by the delta method it follows that

$$(\sqrt{n}(\widehat{F}_{Y|X} - F_{Y|X}), \widehat{\mathbb{G}}) \rightsquigarrow (-\varphi'_{\beta(\cdot)}J^{-1}(\cdot)W(\cdot, \beta(\cdot)), \mathbb{G}) \text{ in } \ell^\infty(\mathcal{Y}\mathcal{X}) \times \ell^\infty(\mathcal{F}), \qquad (\text{E.3})$$

$$(\sqrt{n}(\widehat{F}^*_{Y|X} - \widehat{F}_{Y|X}), \widehat{\mathbb{G}}^*) \rightsquigarrow_\mathbb{P} (-\varphi'_{\beta(\cdot)}J^{-1}(\cdot)W(\cdot, \beta(\cdot)), \mathbb{G}) \text{ in } \ell^\infty(\mathcal{Y}\mathcal{X}) \times \ell^\infty(\mathcal{F}). \qquad (\text{E.4})$$

STEP 3. (Auxiliary: Donskerness). We verify that $\{\varphi_{y,\beta}(Y,X) : (y, \beta) \in \mathcal{Y} \times \mathbb{R}^{d_x}\}$ is P-Donsker with a square integrable envelope. The function classes $\mathcal{F}_1 = \{X'\beta : \beta \in \mathbb{R}^{d_x}\}$, $\mathcal{F}_2 = \{1(Y \leq y) : y \in \mathcal{Y}\}$, and $\{X_q : q = 1, ..., d_x\}$, where q indexes elements of vector X, are VC classes of functions. The final class $\mathcal{G} = \{(\Lambda(\mathcal{F}_1) - \mathcal{F}_2)H(\mathcal{F}_1)X_q : q = 1, ..., d_x\}$ is a Lipschitz transformation of VC classes with Lipschitz coefficient bounded by $\text{const}\|X\|$ and envelope function $\text{const}\|X\|$, which is square-integrable. Hence $\mathcal{G}$ is Donsker by Example 19.9 in van der Vaart (1998). Finally, the map $(\beta, y) \mapsto (\Lambda(X'\beta) - 1\{Y \leq y\})H(X'\beta)X$ is continuous at each $(\beta, y) \in \mathbb{R}^{d_x} \times \mathcal{Y}$ with probability one by the absolute continuity of the conditional distribution of Y (when $\mathcal{Y}$ is not finite).

STEP 4. (Auxiliary: Verification of Conditions of Lemma E.1). We verify conditions (a)-(c) of Lemma E.1. Conditions (a) and (b) are immediate by the assumptions. To verify (c), a straightforward computation gives that for (b, y) in the neighborhood of $(\beta(y), y)$, $\frac{\partial}{\partial(b', y)}\Psi(b, y) = [J(b,y), R(b,y)]$, where, for $H(z) = \lambda(z)/\{\Lambda(z)[1 - \Lambda(z)]\}$ and $h(z) = dH(z)/dz$,

$$J(b,y) := E\left[\{h(X'b)[\Lambda(X'b) - 1(Y \leq y)] + H(X'b)\lambda(X'b)\}XX'\right],$$

and $R(b,y) = -E\left[H(X'b)f_{Y|X}(y|X)X\right]$. Both terms are continuous in (b, y) at $(\beta(y), y)$ for each $y \in \mathcal{Y}$. The computation above as well as verification of continuity follows from using the dominated convergence theorem, and the following ingredients: (1) a.s. continuity of the map $(b, y) \mapsto \frac{\partial}{\partial b'}\varphi_{b,y}(Y, X)$, (2) domination of $\|\frac{\partial}{\partial b'}\varphi_{b,y}(X, Y)\|$ by a square-integrable function

43

const$\|X\|$, (3) a.s. continuity and uniform boundedness of the conditional density function $y \mapsto f_{Y|X}(y|X)$, and (4) $H(X'b)$ bounded uniformly on $b \in \mathbb{R}^{d_x}$, a.s. By assumption $J(y) = J(\beta(y), y)$ is positive-definite uniformly in $y \in \mathcal{Y}$. $\qquad\qquad\square$

References

[1] Aaberge, R., Bjerve, S., and K. Doksum (2005): "Decomposition of rank-dependent measures of inequality by subgroups," *Metron - International Journal of Statistics* vol. 0(3), pp. 493–503.

[2] Abadie, A. (1997): "Changes in Spanish Labor Income Structure during the 1980's: A Quantile Regression Approach," *Investigaciones Economicas* XXI, pp. 253-272.

[3] Abadie, A. (2002):"Bootstrap tests for distributional treatment effects in instrumental variable models," Journal of the American Statistical Association 457, pp. 284–292.

[4] Abbring, J. H., and J. J. Heckman (2007): "Econometric Evaluation of Social Programs, Part III: Distributional Treatment Effects, Dynamic Treatment Effects, Dynamic Discrete Choice, and General Equilibrium Policy Evaluation," in: J.J. Heckman & E.E. Leamer (ed.), *Handbook of Econometrics,* edition 1, volume 6, chapter 72, Elsevier.

[5] Anderson, G. (1996): "Nonparametric Tests of Stochastic Dominance in Income Distribution," *Econometrica* 64, pp. 1183-119.

[6] Angrist, J., Chernozhukov, V., and I. Fernández-Val (2006): "Quantile Regression under Misspecification, with an Application to the U.S. Wage Structure," *Econometrica* 74, pp. 539–563.

[7] Angrist, J., and J.-S. Pischke (2008): *Mostly Harmless Econometrics: An Empiricist's Companion,* Princeton Univesity Press, Princeton.

[8] Autor, D., Katz, L., and M. Kearney (2006a): "The Polarization of the U.S. Labor Market," *American Economic Review* 96, pp. 189–194.

[9] Autor, D., Katz, L., and M. Kearney (2006b): "Rising Wage Inequality: The Role of Composition and Prices," NBER Working Paper w11986.

[10] Autor, D., Katz, L., and M. Kearney (2008): "Trends in U.S. Wage Inequality: Revising the Revisionists," *Review of Economics and Statistics* 90, pp. 300–323.

[11] Barrett, G., and S. Donald (2003): "Consistent Tests for Stochastic Dominance," *Econometrica*, 71, pp. 71–104.

[12] Barrett, G., and S. Donald (2009): "Statistical Inference with Generalized Gini Indexes of Inequality, Poverty, and Welfare," *Journal of Business and Economic Statistics* 27, pp. 1–17.

[13] Bhattacharya, D. (2007): "Inference on Inequality from Household Survey Data," *Journal of Econometrics* 137(2), pp. 674–707.

[14] Blinder, A. (1973): "Wage Discrimination: Reduced Form and Structural Estimates," *The Journal of Human Resources* 8(4), pp. 436-455.

[15] Bollen K.A. and J. Pearl (2012), "Eight Myths about Causality and Structural Equations Models,"to appear in *Handbook of Causal Analysis for Social Research*, S. Morgan (Ed.), Springer.

[16] Buchinsky, M. (1994): "Changes in the US Wage Structure 1963-1987: Application of Quantile Regression," *Econometrica* 62, pp. 405–458.

[17] Burr, B., and H. Doss (1993): "Confidence Bands for the Median Survival Time as a Function of the Covariates in the Cox Model," *Journal of the American Statistical Association* 88, pp. 1330–1340.

[18] Chamberlain, G. (1994): "Quantile Regression, Censoring, and the Structure of Wages," in C. A. Sims (ed.), *Advances in Econometrics, Sixth World Congress,* Volume 1, Cambridge University Press, Cambridge.

[19] Chernozhukov, V. and S. Du (2008): "Extremal Quantiles and Value-at-Risk," *The New Palgrave Dictionary of Economics,* Second Edition. Eds. S. N. Durlauf and L. E. Blume. Palgrave Macmillan.

[20] Chernozhukov, V., and I. Fernández-Val (2005): "Subsampling Inference on Quantile Regression Processes," *Sankhyā* 67, pp. 253–276.

[21] Chernozhukov, V., I. Fernandez-Val and A. Galichon (2010): "Quantile and Probability Curves without Crossing," *Econometrica.*

[22] Chernozhukov, V., I. Fernandez-Val and A. Kowalski (2011): "Quantile Regression with Censoring and Endogeneity," CEMMAP working paper CWP20/11.

[23] Chernozhukov, V., I. Fernandez-Val and B. Melly (2009): "Inference on Counterfactual Distributions," ArXiv: 0904.0951.

[24] Chernozhukov, V., I. Fernandez-Val and B. Melly (2012): "Supplemental Material to Inference on Counterfactual Distributions," unpublished manuscript, MIT.

[25] Chernozhukov, V., and C. Hansen (2005): "An IV Model of Quantile Treatment Effects," *Econometrica* 73, pp. 245–261.

[26] Chernozhukov, V., and C. Hansen (2006): "Instrumental quantile regression inference for structural and treatment effect models," *Journal of Econometrics* 132, pp. 491-525.

[27] Chernozhukov, V., and H. Hong (2002): "Three-step censored quantile regression and extramarital affairs," *Journal of the American Statistical Association* 97, pp. 872–882.

[28] Cox, D. R. (1972): "Regression Models and Life Tables," (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 34, pp. 187–220.

[29] Crump, R. K., Hotz, V. J., Imbens, G.W., and O. A. Mitnik (2009): "Dealing with limited overlap in estimation of average treatment effects," *Biometrika* 96(1), pp. 187-199.

[30] Davidson, R., and J. Duclos (2000): "Statistical inference for stochastic dominance and for the measurement of poverty and inequality," *Econometrica* 68 (6), pp. 1435–64.

[31] DiNardo, J., Fortin, N., and T. Lemieux (1996): "Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach," *Econometrica* 64, pp. 1001–1044.

[32] Donald, S. G., Green, A. A., and H. J. 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, pp. 609–633.

[33] Donald, S. G., and Y.-C. Hsu (2012): "Estimation and Inference for Distribution Functions and Quantile Functions in Treatment Effects Models," unpublished manuscript, UT at Austin.

[34] Doss, H., and R. D. Gill (1992): "An Elementary Approach to Weak Convergence for Quantile Processes, With an Applications to Censored Survival Data," *Journal of the American Statistical Association* 87, pp. 869–877.

[35] Donald, S.G., and Y. Hsu (2012): "Estimation and Inference for Distribution Functions and Quantile Functions in Treatment Effect Models," unpublished manuscript.

[36] Dudley, R. M. (1987): "Universal Donsker Classes and Metric Entropy," *The Annals of Probability* Vol. 15, No. 4, pp. 1306–1326.

[37] Feng, X., He, X., and J. Hu (2011): "Wild bootstrap for quantile regression," *Biometrika* 98, pp. 995–999.

[38] Firpo, S. (2007): "Efficient semiparametric estimation of quantile treatment effects," *Econometrica* 75, pp. 259–276.

[39] Firpo, S., N. Fortin, and T. Lemieux (2009): "Unconditional Quantile Regressions," *Econometrica 77, pp. 953-973.*

[40] Foresi, S., and F. Peracchi (1995): "The Conditional Distribution of Excess Returns: an Empirical Analysis," *Journal of the American Statistical Association* 90, pp. 451-466.

[41] Fortin, N., Lemieux, T., and S. Firpo (2011): "Decomposition Methods in Economics," in: O. Ashenfelter and D. Card (ed.), *Handbook of Labor Economics*, volume 4A, chapter 1, Elsevier.

[42] Giné, E. and R. Nickl (2008): "Uniform central limit theorems for kernel density estimators," *Probability Theory and Related Fields* 141, pp. 333–387.

[43] Gosling, A., S. Machin, and C. Meghir (2000): "The Changing Distribution of Male Wages in the U.K.," *Review of Economic Studies* 67, pp. 635-666.

[44] Hahn, J. (1995), "Bootstrapping Quantile Regression Estimators," *Econometric Theory* 11, pp. 105–121.

[45] Han, A., and J. A. Hausman (1990), "Flexible Parametric Estimation of Duration and Competing Risk Models," *Journal of Applied Econometrics* 5, pp. 1–28.

[46] Heckman, J. J. (1998), "Detecting discrimination", *The Journal of Economic Perspectives* 12, pp. 101-116.

[47] Heckman, J. J. (2008), "Econometric Causality", *International Statistical Review* 76, pp. 1–27.

[48] Heckman, J.J., Ichimura, H., Smith, J., and P. Todd (1998), "Characterizing Selection Bias Using Experimental Data," *Econometrica* 66(5), pp. 1017-1098.

[49] Heckman, J., R. LaLonde, and J. Smith (1999): "The economics and econometrics of active labor market programs," in: O. Ashenfelter and D. Card (ed.), *Handbook of Labor Economics*, volume 3A, chapter 31, Elsevier.

[50] Heckman, J., and R. Robb, (1985),: "Alternative Methods for Evaluating the Impact of Interventions," in Heckman and Singer (eds.), Longitudinal Analysis of Labor Market Data, Cambridge, Cambridge University Press.

[51] Heckman, J. J., Smith, J., and N. Clements (1997), "Making the Most Out of Programme Evaluations and Social Experiments: Accounting for Heterogeneity in Programme Impacts," *The Review of Economic Studies,* Vol. 64, pp. 487–535.

[52] Heckman J. J. and E. Vytlacil (2005): "Structural equations, treatment effects, and econometric policy evaluation," *Econometrica* 73, pp. 669-738.

[53] Heckman J. J. and E. Vytlacil (2007): "Econometric Evaluation of Social Programs, Part I: Causal Models, Structural Models and Econometric Policy Evaluation," in: J.J. Heckman & E.E. Leamer (ed.), *Handbook of Econometrics*, edition 1, volume 6, chapter 70, Elsevier.

[54] Hirano, K., Imbens, G. W., and G. Ridder, (2003), "Efficient estimation of average treatment effects using the estimated propensity score," *Econometrica,* Vol. 71, pp. 1161–1189.

[55] Holland, P. W. (1986), "Statistics and Causal Inference," *Journal of the American Statistical Association* 81, pp. 945-960.

[56] Horvitz, D., and D. Thompson (1952), "A Generalization of Sampling Without Replacement from a Finite Universe," *Journal of the American Statistical Association* 47, pp. 663–685.

[57] Imbens, G. W. (2004): "Nonparametric Estimation of Average Treatment Effects Under Exogeneity: A Review," *The Review of Economics and Statistics* 86, pp. 4–29.

[58] Kato, K. (2011), "A note on moment convergence of bootstrap M-estimators," *Statistics & Decisions* 28, pp. 51-61.

[59] Klecan, L., McFadden, R. and D. McFadden (1991): "A Robust Test for Stochastic Dominance," unpublished manuscript, MIT.

[60] Koenker, R. (2005), *Quantile Regression.* Econometric Society Monograph Series 38, Cambridge University Press.

[61] Koenker R. and G. Bassett (1978): "Regression Quantiles," *Econometrica* 46, pp. 33–50.

[62] Koenker, R. and J. Yoon (2009): "Parametric links for binary choice models: A Fisherian-Bayesian colloquy," *Journal of Econometrics* 152(2), pp. 120–130.

[63] Koenker, R., and Z. Xiao (2002): "Inference on the Quantile Regression Process," *Econometrica* 70, no. 4, pp. 1583–1612.

[64] Lemieux, T. (2006): "Increasing Residual Wage Inequality: Composition Effects, Noisy Data, or Rising Demand for Skill?," *American Economic Review* 96, pp. 451–498.

[65] Linton, O., Maasomi, E., and Y. Whang (2005): "Testing for Stochastic Dominance under general conditions: A subsampling approach," *Review of Economic Studies* 72, pp. 735–765.

[66] Linton, O., Song, K., and Y. Whang (2010): "An improved bootstrap test of stochastic dominance," *Journal of Econometrics* 154, pp. 186-202.

[67] Machado, J., and J. Mata (2005): "Counterfactual Decomposition of Changes in Wage Distributions Using Quantile Regression," *Journal of Applied Econometrics* 20, pp. 445–465.

[68] Maier, M. (2011): "Tests for distributional treatment effects under unconfoundedness," *Economics Letters* 110, pp. 49-51.

[69] McFadden, D. (1989): "Testing for Stochastic Dominance," in Part II of T. Fomby and T. K. Seo (eds) *Studies in the Economics of Uncertainty* (in honor of J. Hadar), Springer-Verlag.

[70] Oaxaca, R. (1973): "Male-Female Wage Differentials in Urban Labor Markets," *International Economic Review* 14(3), pp. 693-709.

[71] Powell, J. L. (1986): "Censored Regression Quantiles," *Journal of Econometrics* 32, pp. 143-155.

[72] Radulović, D. and M. Wegkamp (2003): "Necessary and sufficient conditions for weak convergence of smoothed empirical processes," *Statistics & Probability Letters* 61, pp. 321–336.

[73] Resnick, S. I. (1987), *Extreme values, regular variation, and point processes*, Applied Probability. A Series of the Applied Probability Trust, 4. Springer-Verlag, New York.

[74] Rosenbaum, P., and D. Rubin, (1983): "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika* 70, pp. 41–55.

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

[76] Rothe, C. (2012): "Partial Distributional Policy Effects," *Econometrica* 80, pp. 2269-2301.

[77] Rothe, C., and D. Wied (2012): "Specification Testing in a Class of Conditional Distributional Models," *Journal of the American Statistical Association*, forthcoming.

[78] Rubin, D. B. (1978): "Bayesian Inference for Causal Effects: The Role of Randomization," *The Annals of Statistics* 6, pp. 34-58.

[79] Stock, J. H. (1989): "Nonparametric Policy Analysis," *Journal of the American Statistical Association* 84, pp. 567-575.

[80] Stock, J. H. (1991): "Nonparametric Policy Analysis: An Application to Estimating Hazardous Waste Cleanup Benefits," in *Nonparametric and Semiparametric Methods in Econometrics and Statistics,* eds. W. Barnett, J. Powell, and G. Tauchen, Cambridge, U.K.: Cambridge University Press.

[81] van der Vaart, A. (1998): *Asymptotic statistics,* Cambridge Series in Statistical and Probabilistic Mathematics, 3.

[82] van der Vaart, A., and J. Wellner (1996): *Weak convergence and empirical processes: with applications to statistics*, New York: Springer.
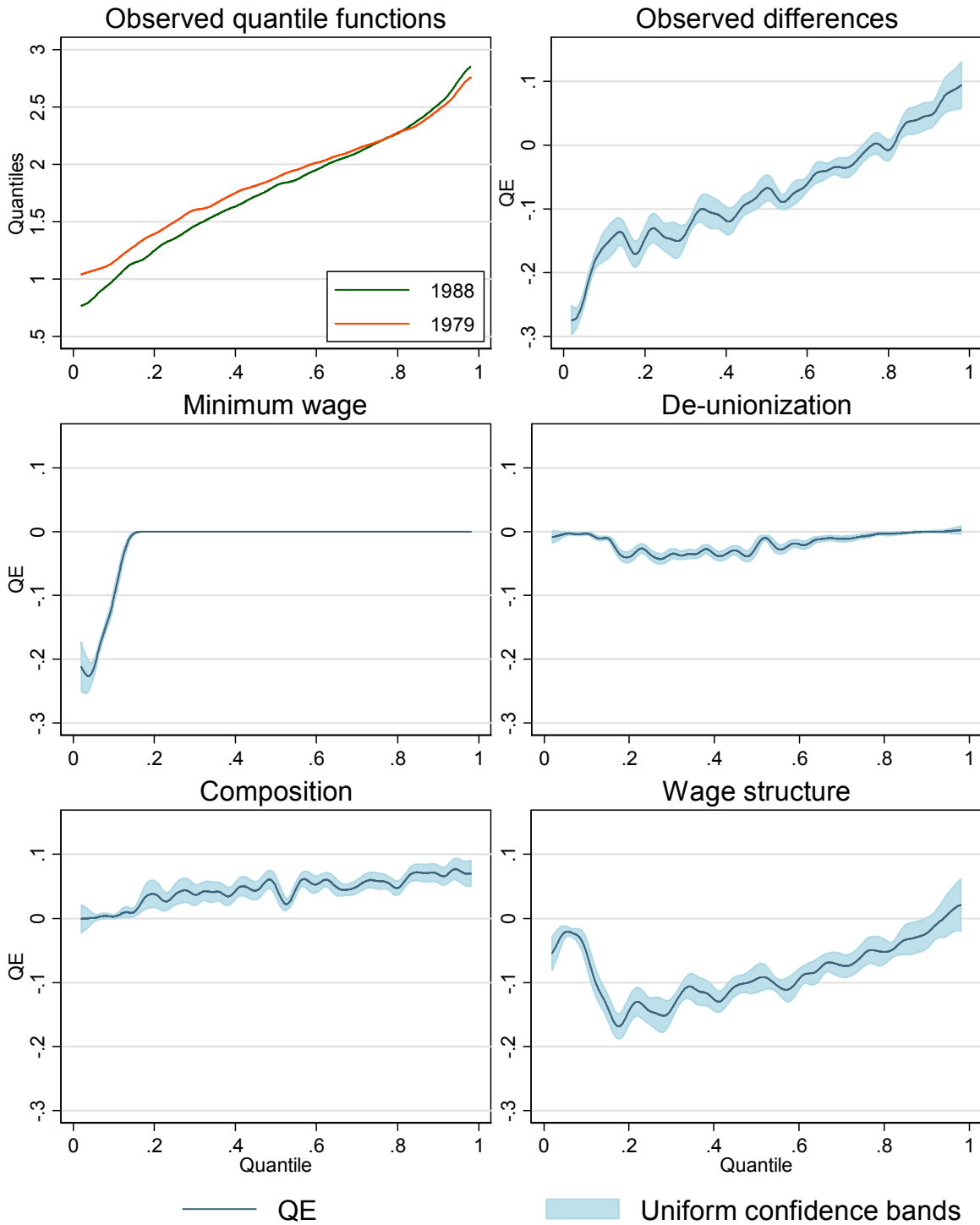
Figure 1. Observed quantile functions, observed differences between the quantile functions and their decomposition into four quantile effects. The 95% simultaneous confidence bands were obtained by empirical bootstrap with 100 repetitions.
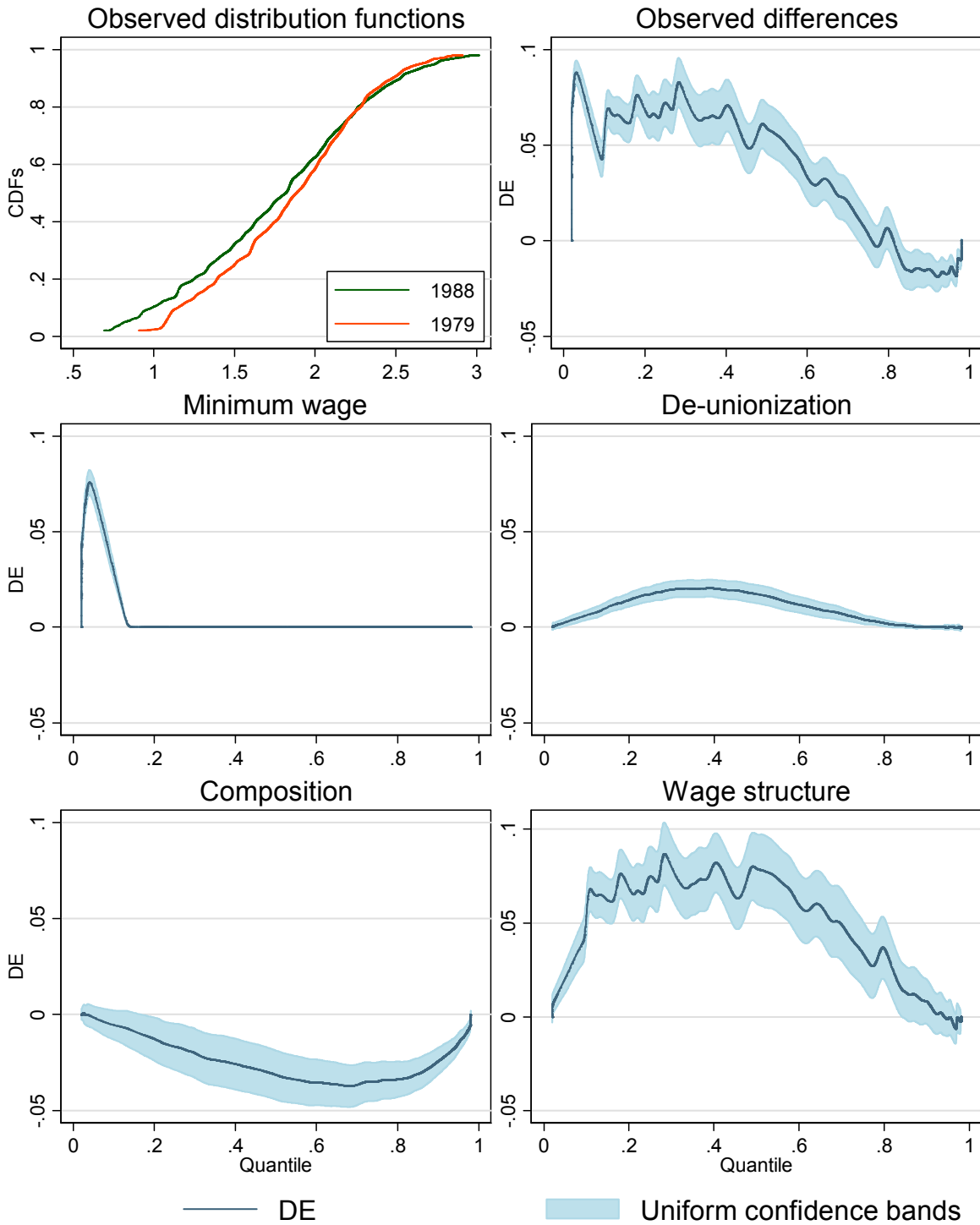
Figure 2. Observed distribution functions, observed differences between the distribution functions and their decomposition into four distribution effects. The 95% simultaneous confidence bands were obtained by empirical bootstrap with 100 repetitions.
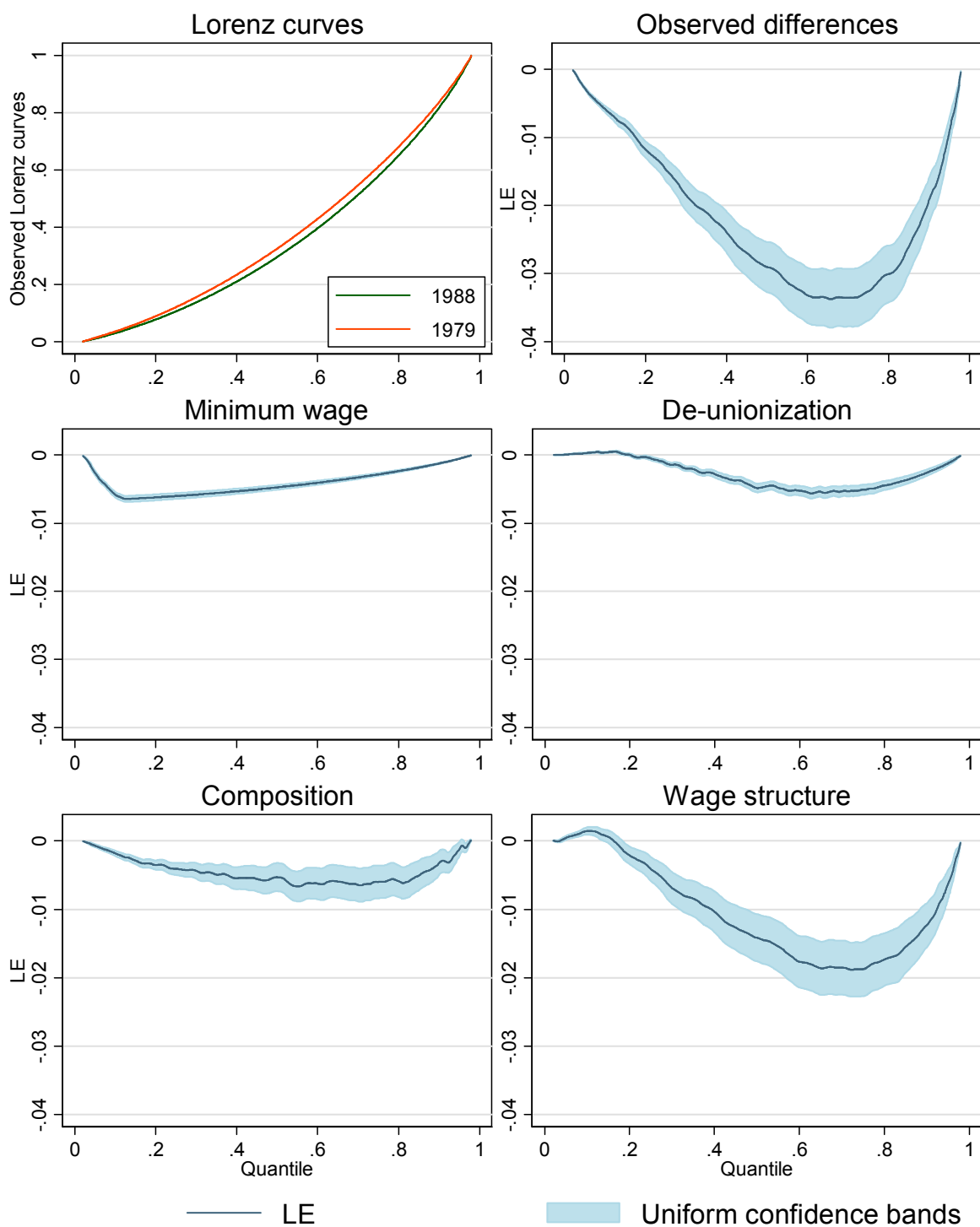
Figure 3. Observed Lorenz curves, observed differences between the Lorenz curves and their decomposition into four Lorenz effects. The 95% simultaneous confidence bands were obtained by empirical bootstrap with 100 repetitions.

Table 1: Decomposing changes in measures of wage dispersion

Statistic	Total change	Effect of: Minimum wage	Unions	Composition	Wage structure
Standard Deviation	8.0 (0.3)	2.8 (0.1)	0.7 (0.0)	1.8 (0.2)	2.7 (0.3)
		35.4 (1.4)	8.5 (0.6)	22.9 (1.9)	33.1 (2.4)
90-10	21.5 (1.0)	11.2 (0.1)	0.0 (0.0)	9.2 (0.8)	1.1 (1.3)
		52.1 (2.4)	0.0 (0.1)	42.6 (4.4)	5.3 (5.9)
50-10	11.3 (1.4)	11.2 (0.1)	-2.0 (1.0)	5.1 (0.4)	-3.1 (1.1)
		99.6 (14.1)	-17.9 (11.2)	45.5 (8.3)	-27.2 (14.0)
90-50	10.2 (1.2)	0.0 (0.0)	2.0 (1.0)	4.0 (0.8)	4.2 (1.1)
		0.0 (0.0)	19.7 (8.4)	39.3 (8.8)	41.0 (9.8)
75-25	15.4 (1.1)	0.0 (0.0)	4.1 (1.0)	0.3 (1.3)	11.1 (1.2)
		0.0 (0.0)	26.5 (6.2)	1.7 (8.6)	71.8 (8.7)
95-5	33.0 (2.1)	23.0 (0.7)	0.0 (0.6)	8.5 (1.1)	1.4 (1.5)
		69.9 (4.1)	0.0 (1.7)	25.8 (2.6)	4.3 (4.4)
Gini coefficient	4.1 (0.1)	1.3 (0.0)	0.5 (0.0)	0.3 (0.1)	2.0 (0.1)
		32.1 (1.2)	11.7 (0.6)	6.8 (1.8)	49.4 (1.8)

Notes: The logit distribution regression model has been applied. All numbers are in %. Bootstrapped standard errors with 100 repetitions are given in parenthesis. The second line in each cell indicates the percentage of total variation.

SUPPLEMENTAL MATERIAL TO
INFERENCE ON COUNTERFACTUAL DISTRIBUTIONS

VICTOR CHERNOZHUKOV[†]   IVÁN FERNÁNDEZ-VAL[§]   BLAISE MELLY[‡]

ABSTRACT. This supplemental material contains four appendices to Chernozhukov, Fernández-Val and Melly (2012). In Appendix A, we show the validity of bootstrap confidence bands for general counterfactual functionals. In Appendix B, we compare the performance of quantile and distribution regression as estimators of the conditional and counterfactual distribution functions in Monte Carlo simulations. We calibrate our data generating process to fit several characteristics of the Current Population Survey data used in Section 6 of the paper. We find that quantile regression works better than distribution regression when the distribution of the dependent variable is continuous but it performs worse in the presence of a small and realistic amount of rounding in the data. Appendices C and D complement the empirical results presented in Section 6 of the paper. Appendix C contains decriptive statistics, additional results for men and all the results for women. Appendix D gives more details about the variance decomposition into between-group and within-group components.

Key Words: counterfactual distribution, quantile regression, distribution regression, minimum wage, unions

APPENDIX A. VALIDITY OF BOOTSTRAP CONFIDENCE BANDS AND UNIFORM CONSISTENCY
OF BOOSTRAP VARIANCE ESTIMATORS

We consider the general counterfactual functionals

$$\Delta_t(w) = \phi(F_{Y\langle j|k\rangle} : (j,k) \in \mathcal{JK})(w),$$

and their plug-in estimators

$$\widehat{\Delta}_t(w) = \phi(\widehat{F}_{Y\langle j|k\rangle} : (j,k) \in \mathcal{JK})(w),$$

for $w \in \mathcal{W}$. In Sections 4 and 5 of Chernozhukov, Fernández-Val and Melly (2012, CFM in the following), we established conditions for the functional central limit theorem:

$$\widehat{Z} := \sqrt{n}(\widehat{\Delta}_t - \Delta_t) \rightsquigarrow Z \text{ in } \ell^\infty(\mathcal{W}),$$

where Z is a zero-mean Gaussian process with a.s. continuous sample paths, and pointwise variance function $\Sigma(w) = E[Z^2(w)]$. Let

$$\widehat{\Delta}_t^*(w) = \phi(\widehat{F}_{Y\langle j|k\rangle}^* : (j,k) \in \mathcal{JK})(w),$$

be the bootstrap version of $\widehat{\Delta}_t(w)$. We also gave conditions for the bootstrap functional central limit theorem:

$$\widehat{Z}^* := \sqrt{n}(\widehat{\Delta}_t^* - \widehat{\Delta}_t) \rightsquigarrow_\mathbb{P} Z \text{ in } \ell^\infty(\mathcal{W}).$$

Let $\widehat{t}_{1-\alpha}^*$ be the $1-\alpha$ quantile of the bootstrap version of the Kolmogorov-Smirnov maximal t-statistic:

$$\widehat{t}^* = \sup_{w \in \mathcal{W}} \widehat{\Sigma}^*(w)^{-1/2}|\widehat{Z}^*(w)|,$$

conditional on the data. As a robust estimator of $\Sigma(w)$, we consider the rescaled squared bootstrap quantile spread,

$$\widehat{\Sigma}^*(w) := [\widehat{Z}_{\tau_1}^*(w) - \widehat{Z}_{\tau_2}^*(w)]^2/[Z_{\tau_1} - Z_{\tau_2}]^2, \tag{A.1}$$

where $\widehat{Z}_\tau^*(w)$ is the τ quantile of $\widehat{Z}^*(w)$ conditional of the data, Z_τ is the τ quantile of the $N(0,1)$, and $0 < \tau_2 < \tau_1 < 1$. The bootstrap versions of the end-point functions of the $1-\alpha$ confidence band are

$$\widehat{\Delta}_t^{\pm *}(w) = \widehat{\Delta}_t(w) \pm \widehat{t}_{1-\alpha}^* \widehat{\Sigma}^*(w)^{1/2}/\sqrt{n}.$$

The following lemma establishes that the bootstrap confidence band $[\widehat{\Delta}_t^{-*}(w), \widehat{\Delta}_t^{+*}(w)]$ covers $\Delta_t(w)$ uniformly over $w \in \mathcal{W}$ with asymptotic probability $1-\alpha$. It is based on Theorems 1 and 2 and Lemma 1 in Chernozhukov and Fernandez-Val (2005), adapted to our setting.

Lemma A.1 (Consistency of Bootstrap Confidence Bands). *Suppose that $\widehat{Z} \rightsquigarrow Z$ and $\widehat{Z}^* \rightsquigarrow_\mathbb{P} Z$ in $\ell^\infty(\mathcal{W})$, where Z is a tight zero-mean Gaussian process with continuos paths in $\mathcal{W}$ and variance*

function $\Sigma(w)$ that is bounded away from zero and from above uniformly over $w \in \mathcal{W}$. Then, (a) the variance estimator specified in (A.1) is uniformly consistent:

$$\sup_{w \in \mathcal{W}} |\widehat{\Sigma}^*(w) - \Sigma(w)| \to_\mathbb{P} 0;$$

and (b)

$$\lim_{n \to \infty} \mathbb{P} \left\{ \Delta_t(w) \in [\widehat{\Delta}_t^{-*}(w), \widehat{\Delta}_t^{+*}(w)] \text{ for all } w \in \mathcal{W} \right\} = 1 - \alpha.$$

In order to state the proof formally, we follow the notation and definitions in van der Vaart and Wellner (1996). Let D_n denote the data vector and M_n be the vector of random variables used to generate bootstrap draws or simulation draws given D_n (this may depend on the particular resampling or simulation method). Consider the random element $\mathbb{Z}_n^* = \mathbb{Z}_n(D_n, M_n)$ in a normed space $\mathbb{D}$. We say that the bootstrap law of $\mathbb{Z}_n^*$ consistently estimates the law of some tight random element $\mathbb{Z}$ and write $\mathbb{Z}_n^* \rightsquigarrow_\mathbb{P} \mathbb{Z}$ in $\mathbb{D}$ if

$$d_{\text{BL}}(\mathbb{Z}_n^*, \mathbb{Z}) := \sup_{h \in \text{BL}_1(\mathbb{D})} |E_{M_n} h(\mathbb{Z}_n^*) - Eh(\mathbb{Z})| \to_\mathbb{P} 0, \tag{A.2}$$

where $\text{BL}_1(\mathbb{D})$ denotes the space of functions mapping $\mathbb{D}$ to the real line with Lipschitz norm at most 1 and E_{M_n} denotes the conditional expectation with respect to M_n given the data D_n.

Proof of Lemma A.1 Step (1). Here we establish claim (a). Note that for each $w \in \mathcal{W}$, $Z(w) \sim N(0, \Sigma(w))$, so that the τ quantile of $Z(w)$ is $Z_\tau(w) = \Sigma(w)^{1/2} Z_\tau$, and $\Sigma(w) = [Z_{\tau_1}(w) - Z_{\tau_2}(w)]^2 / [Z_{\tau_1} - Z_{\tau_2}]^2$. Let

$$F(\nu, w) := P[Z(w) \le \nu] = \Phi(\Sigma^{-1/2}(w)\nu),$$

where Φ is the cdf of the $N(0,1)$. Note that $w \mapsto \Sigma(w)^{1/2}$ is continuous in $w \in \mathcal{W}$ by Gaussianity of Z and a.s. continuity of its sample paths with respect to $w \in \mathcal{W}$. Moreover, $w \mapsto \Sigma^{-1/2}(w)$ is uniformly continuous on $\mathcal{W}$ due to continuity of $w \mapsto \Sigma^{1/2}(w)$ and $\Sigma^{1/2}(w)$ being bounded away from zero uniformly over $\mathcal{W}$. This implies that $(\nu, w) \mapsto F(\nu, w)$ is uniformly continuous on $(\nu, w) \in \mathbb{R} \times \mathcal{W}$, and that $Z_\tau(w) = \inf_{\nu \in \mathbb{R}} \{F(\nu, w) \ge \tau\}$ is uniformly continuous in $(\tau, w) \in \mathcal{T} \times \mathcal{W}$, where $\mathcal{T}$ is a compact subinterval of $(0, 1)$. In the paragraph below we shall deduce that $\widehat{F}(\nu, w) := E_{M_n} 1\{\widehat{Z}^*(w) \le \nu\}$ obeys

$$\widehat{F}(\nu, w) \to_\mathbb{P} F(\nu, w), \text{ uniformly in } (\nu, w) \in \mathbb{R} \times \mathcal{W}. \tag{A.3}$$

By Lemma 1 in Chernozhukov and Fernandez-Val (2005), these properties imply that

$$\widehat{Z}_\tau^*(w) \to_\mathbb{P} Z_\tau(w), \text{ uniformly in } (\tau, w) \in \mathcal{T} \times \mathcal{W}. \tag{A.4}$$

Therefore uniformly in $w \in \mathcal{W}$, by the continuous mapping theorem,

$$\widehat{\Sigma}^*(w) = \frac{[\widehat{Z}_{\tau_1}^*(w) - \widehat{Z}_{\tau_2}^*(w)]^2}{[Z_{\tau_1} - Z_{\tau_2}]^2} \to_\mathbb{P} \frac{[Z_{\tau_1}(w) - Z_{\tau_2}(w)]^2}{[Z_{\tau_1} - Z_{\tau_2}]^2} = \Sigma(w).$$

The argument to show (A.3) is based in part on the proof of Theorem 2 in Chernozhukov and Fernandez-Val (2005). Let $u \mapsto K_{\delta,\nu}(u) := 1(u \leq \nu) + 1(\nu < u \leq \nu + \delta)(\nu + \delta - u)/\delta$ denote a smoothed approximation to the indicator function $u \mapsto 1(u \leq \nu)$, with $\delta > 0$ denoting the smoothing parameter. Then, the collection of maps mapping $z \in \ell^{\infty}(\mathcal{W})$ to $\mathbb{R}$, defined as: $\mathcal{G}_{\delta} = \{z \mapsto K_{\delta,\nu}(z(w)) : \nu \in \mathbb{R}, w \in \mathcal{W}\}$, is a subset of $\delta^{-1}\mathrm{BL}_1(\ell^{\infty}(\mathcal{W}))$. By $\widehat{Z}^* \rightsquigarrow_{\mathbb{P}} Z$ in $\ell^{\infty}(\mathcal{W})$, $\sup_{g \in \mathcal{G}_{\delta}} |E_{M_n} g(\widehat{Z}^*) - Eg(Z)| \leq \delta^{-1} d_{BL}(\widehat{Z}^*, Z) \rightarrow_{\mathbb{P}} 0$. Therefore, uniformly in $w \in \mathcal{W}$ and $\nu \in \mathbb{R}$,

$$\widehat{F}(\nu, w) = E_{M_n} 1\{\widehat{Z}^*(w) \leq \nu\} \leq E_{M_n} K_{\delta,\nu}[\widehat{Z}^*(w)] \rightarrow_{\mathbb{P}} EK_{\delta,\nu}[Z(w)]. \tag{A.5}$$

We have that $F(\nu, w) \leq EK_{\delta,\nu}[Z(w)] \leq F(\nu + \delta, w)$. By uniform continuity of $F(\cdot, \cdot)$, as $\delta \searrow 0$, $EK_{\delta,\nu}[\widehat{Z}(w)] \rightarrow F(\nu, w)$, uniformly in $(\nu, w) \in \mathbb{R} \times \mathcal{W}$. Therefore, we conclude that $\widehat{F}(\nu, w) \leq F(\nu, w) + o_{\mathbb{P}}(1)$, uniformly in $(\nu, w) \in \mathbb{R} \times \mathcal{W}$. Arguing similarly, we can also deduce that $\widehat{F}(\nu, w) \geq F(\nu, w) + o_{\mathbb{P}}(1)$, uniformly in $(\nu, w) \in \mathbb{R} \times \mathcal{W}$.

Step (2). Here we establish that

$$\widehat{t}^* \rightsquigarrow_{\mathbb{P}} \sup_{w \in \mathcal{W}} \Sigma(w)^{-1/2} |Z(w)| =: t,$$

in $\mathbb{R}$. Let $\widetilde{t}^* = \sup_{w \in \mathcal{W}} \Sigma(w)^{-1/2} |\widehat{Z}^*(w)|$. The collection of functions mapping elements $z \in \ell^{\infty}(\mathcal{W})$ to the real line, defined as $\mathcal{G} = \{z \mapsto g_m(z) := m(\|\Sigma^{-1/2}(\cdot)|z(\cdot)|\|_{\mathcal{W}}) : m \in \mathrm{BL}_1(\mathbb{R})\}$, is a subset of $M \cdot \mathrm{BL}_1(\ell^{\infty}(\mathcal{W}))$, where $M = \sup\{\Sigma^{-1/2}(w) : w \in \mathcal{W}\}$, since $|g_m(z) - g_m(\tilde{z})| \leq M(|z - \tilde{z}|_{\mathcal{W}} \wedge 1)$. Therefore, $d_{BL}(\widetilde{t}^*, t) \leq M d_{BL}(\widehat{Z}^*, Z) \rightarrow_{\mathbb{P}} 0$, where the latter holds by the definition of $\widehat{Z}^* \rightsquigarrow_{\mathbb{P}} Z$. Moreover, $\widehat{t}^* = \widetilde{t}^* + o_{\mathbb{P}}(1)$, since uniformly over $\mathcal{W}$, $\widehat{\Sigma}^*(w)^{-1/2} |\widehat{Z}^*(w)| = \Sigma^*(w)^{-1/2} |\widehat{Z}^*(w)| + [\widehat{\Sigma}^*(w)^{-1/2} - \Sigma(w)^{-1/2}]|\widehat{Z}^*(w)| = \Sigma(w)^{-1/2}|\widehat{Z}^*(w)| + o_{\mathbb{P}}(1)$ by part (a) of the lemma. Hence $\widehat{t}^* \rightsquigarrow_{\mathbb{P}} t$ follows from

$$d_{BL}(\widehat{t}^*, t) \leq d_{BL}(\widetilde{t}^*, t) + E_{M_n}[|\widetilde{t}^* - \widehat{t}^*| \wedge 1] \rightarrow_{\mathbb{P}} 0,$$

where $E_{M_n}[|\widetilde{t}^* - \widehat{t}^*| \wedge 1] \rightarrow_{\mathbb{P}} 0$, by Markov inequality and $E[E_{M_n}[|\widetilde{t}^* - \widehat{t}^*| \wedge 1] \leq E[|o_{\mathbb{P}}(1)| \wedge 1] \rightarrow 0$.

Step (3). Here we show that the limit t is a continuous random variable and $\widehat{t}^*_{1-\alpha} \rightarrow_{\mathbb{P}} t_{1-\alpha}$, where $t_{1-\alpha}$ is the $1 - \alpha$ quantile of t. The continuity of the distribution of t follows from Theorem 11.1 of Davydov, Lifshits, and Smordina (1998) using that $\Sigma(w)$ is non-degenerate over $\mathcal{W}$. The claim $\widehat{t}^*_{1-\alpha} \rightarrow_{\mathbb{P}} t_{1-\alpha}$ follows from Step (2), because by the same argument as in Step (1), $\widehat{t}^* \rightsquigarrow_{\mathbb{P}} t$ in $\mathbb{R}$ and t having a continuous distribution function implies convergence of quantiles, namely $\widehat{t}^*_{1-\alpha} \rightarrow_{\mathbb{P}} t_{1-\alpha}$. We also remark that this part of the proof is standard, see e.g. Beran (1984).

Step (4). Here we show claim (b) of the lemma. The argument is standard (e.g., Beran, 1984). Let $\widehat{t} = \sup_{w \in \mathcal{W}} \widehat{\Sigma}^*(w)^{-1/2} |\widehat{Z}(w)|$. By the continuous mapping theorem $\widehat{t} \rightsquigarrow t$ in $\mathbb{R}$. The event

$$\left\{ \Delta_t(w) \in [\widehat{\Delta}_t^{-*}(w), \widehat{\Delta}_t^{+*}(w)] \text{ for all } w \in \mathcal{W} \right\}$$

is equivalent to the event $\{\widehat{t} \leq \widehat{t}^*_{1-\alpha}\}$. By $\widehat{t}^*_{1-\alpha} \rightarrow_{\mathbb{P}} t_{1-\alpha}$ in part (3), for any $\varepsilon > 0$ and all sufficiently large n, $t_{1-\alpha} - \varepsilon < \widehat{t}^*_{1-\alpha} < t_{1-\alpha} + \varepsilon$. Consequently, $\mathbb{P}\{\widehat{t} \leq t_{1-\alpha} - \varepsilon\} + o(1) \leq \mathbb{P}\{\widehat{t} \leq$

$\widehat{t}^*_{1-\alpha}\} \leq \mathbb{P}\{\widehat{t} \leq t_{1-\alpha} + \varepsilon\} + o(1)$. Taking limits as $n \to \infty$

$$\mathbb{P}\{t \leq t_{1-\alpha} - \varepsilon\} \leq \liminf_{n \to \infty} \mathbb{P}\{\widehat{t} \leq \widehat{t}^*_{1-\alpha}\} \leq \limsup_{n \to \infty} \mathbb{P}\{\widehat{t} \leq \widehat{t}^*_{1-\alpha}\} \leq \mathbb{P}\{t \leq t_{1-\alpha} + \varepsilon\},$$

where we use that $\widehat{t} \rightsquigarrow t$ in $\mathbb{R}$ and continuity of the distribution function of t. The result follows from taking $\varepsilon \to 0$, using the continuity of the distribution of t, and $\mathbb{P}\{t \leq t_{1-\alpha}\} = 1 - \alpha$. $\square$

Appendix B. Comparison of Quantile and Distribution Regression: a Simulation Exercise

B.1. **Data generating processes.** In this appendix we compare quantile and distribution regression as estimators of the conditional and counterfactual distribution functions. We calibrate the data generating processes to fit several characteristics of the CPS data sets used in the application in Section 6 of CFM. In particular, we draw the Monte Carlo samples of X_{88} from the sample distribution in 1988 (containing 74,661 observations). We consider a model with 8 covariates: a dummy variable for living in metropolitan areas, a dummy variable for part-time work, a dummy variable for not being white, experience, experience square, a union indicator, education and education squared. These covariates are only a subset of the covariates that we include in the application because we want also to consider samples of moderate sizes, which would lead to frequent multicollinearity problems with the 45 regressors used in the application. In order to estimate a counterfactual distribution, we also draw independent Monte Carlo samples of X_{79} from the sample distribution in 1979 (containing 21,483 observations).

We consider three different data generating processes (DGP) for the conditional distribution of Y_{88} given X_{88}. We start with a very simple DGP and then show that the conclusions do not change with more realistic models. In this first DGP, the dependent variable is linear in X:

$$Y_i = X_i'\beta + u_i, \ i = 1, ..., n$$

where the errors u_i are iid logistically distributed. The coefficient vector β is set to its estimate obtained from the OLS regression of the log hourly wage on X in the whole 1988 sample. Similarly, the variance of u_i is set to the OLS residual variance.

The second DGP we consider is the same as the first one except that the errors u_i are drawn iid from the empirical distribution of the OLS residuals in the whole sample.

Finally, in the third DGP, we consider a multiplicative linear heteroscedastic model:

$$Y_i = X_i'\beta + (X_i'\delta) \cdot u_i, \ i = 1, ..., n$$

where the errors u_i are randomly drawn from the empirical distribution of the OLS residuals in the whole sample re-scaled to have variance one. The coefficient vector β is the same as in the first two DGPs. The coefficient vector δ is set to its OLS estimate obtained by regressing the squared OLS residuals on X in the whole 1988 sample. (We checked that $X_i'\delta > 0$ for $i = 1, ..., n$.)

Our analysis of the wage distribution in Section 6 reveals a considerable amount of discreteness in the data. A natural mass point is found at the level of the minimum wage but other mass points are observed at different levels. Rounded hourly wages are natural focal points in wage negociation, which could explain the presence of mass points. In order to analyze the effect of rounding on the quality of the estimation, we consider a version of the three DGPs where wages are rounded to the next dollar with probability 1/3.

Figure B1 compares the amount of rounding present in the wage data in 1988 and the amount of rounding implied by our data generating process.[1] It appears that the simulated data matches very well the probability that the wage takes exactly one of the 10 most frequent values. For instance, the observed wages take one of the two most frequent values with probability 8.7% while this probability is 8.6% for the simulated data. Our data generating process does not round enough the dependent variable when we consider the values that are not among the ten most frequent values. In this sense, the discreteness implied by our DGPs is more modest than that observed in the data.

B.2. **Estimands of interest and estimators.** The first estimand of interest is the conditional distribution function of Y given X:

$$F_{Y_{88}|X_{88}}(y|x) \text{ for } (y,x) \in \mathcal{YX},$$

where $\mathcal{YX}$ is the joint support of Y_{88} and X_{88}. We are interested in the conditional distribution because this quantity is often of direct economic interest and it is the first step in the estimation of the counterfactual distributions. Our second and main estimand is the counterfactual distribution obtain by integrating this distribution over the empirical marginal distribution of X in 1979:

$$F_{Y\langle 88|79\rangle}(y) = \int_{\mathcal{X}} F_{Y_{88}|X_{88}}(y|x)\, dF_{X_{79}}(x).$$

We implement the distribution regression estimator using the logistic link function. For 500 different cutoff values y located at equidistant marginal quantiles of Y_{88}, we regress $1(Y_{88} \leq y)$ on X_{88} using a linear logit regression. The estimated conditional distribution is obtained as

$$\widehat{F}_{Y_{88}|X_{88}}^{DR}(y|x) = \Lambda(x'\widehat{\beta}(y)),$$

where $\Lambda(\cdot)$ is the logistic distribution function. Since this estimated conditional distribution function may be non-monotonic in y, we also apply the monotonization method of Chernozhukov, Fernandez-Val and Galichon (2010) based on rearrangement. The rearranged cdf is[2]

$$\widehat{F}_{Y_{88}|X_{88}}^{*}(y|x) = \inf\left\{u : \int_0^\infty 1\left(\widehat{F}_{Y_{88}|X_{88}}^{DR}(t|x) \leq u\right) dt \geq y\right\}.$$

---

[1] We plot these probabilities for the first DGP. The probabilities are very similar for the two other DGPs.

[2] This expression is simplified by the fact that Y is non-negative.

To implement the quantile estimator, we estimate 500 linear quantile regressions of Y_{88} on X_{88} (Koenker and Bassett, 1978) and obtain

$$\widehat{Q}_{Y_{88}|X_{88}}(\tau|x) = x'\widehat{\beta}(\tau), \text{ for } \tau = 0.001, ..., 0.999.$$

We invert this estimated quantile function to obtain the cdf using

$$\widehat{F}^{QR}_{Y_{88}|X_{88}}(y|x) = \frac{1}{1000} + \int_{1/1000}^{999/1000} 1\left(\widehat{Q}_{Y_{88}|X_{88}}(u|x) \leq y\right) du.$$

The counterfactual distribution of interest, $F_{Y\langle 88|79\rangle}(y)$, is estimated by averaging the estimated conditional distribution in 1988 over the covariate distribution in 1979. The estimates of this counterfactual distribution are all obtained as

$$\widehat{F}_{Y\langle 88|79\rangle}(y) = \frac{1}{n}\sum_{i=1}^{n}\widehat{F}_{Y_{88}|X_{88}}(y|x_{79,i})$$

where $\widehat{F}_{Y_{88}|X_{88}}(y|x)$ is $\widehat{F}^{DR}_{Y_{88}|X_{88}}(y|x)$, $\widehat{F}^{*}_{Y_{88}|X_{88}}(y|x)$ or $\widehat{F}^{QR}_{Y_{88}|X_{88}}(y|x)$. Thus, two independent samples of size n are used to estimate this counterfactual distribution: n observations drawn from (X_{88}, Y_{88}) to obtain $\widehat{F}_{Y_{88}|X_{88}}(y|x)$ and n observations drawn from X_{79} to obtain $(x_{79,1}, \ldots, x_{79,n})$

B.3. **Results.** We measure the performance of the estimators by the integrated mean squared error (MSE) and the integrated Anderson-Darling weighted MSE. The MSEs for the counterfactual and conditional distributions are

$$\int_{\mathcal{Y}}\left(\widehat{F}_{Y\langle 88|79\rangle}(y) - F_{Y\langle 88|79\rangle}(y)\right)^2 dF_{Y\langle 88|79\rangle}(y),$$

$$\int_{\mathcal{X}}\int_{\mathcal{Y}}\left(\widehat{F}_{Y_{88}|X_{88}}(y|x) - F_{Y_{88}|X_{88}}(y|x)\right)^2 dF_{Y_{88}|X_{88}}(y|x)\,dF_{X_{88}}(x).$$

Similarly, the Anderson-Darling weighted MSEs are

$$\int_{\mathcal{Y}}\frac{\left(\widehat{F}_{Y\langle 88|79\rangle}(y) - F_{Y\langle 88|79\rangle}(y)\right)^2}{F_{Y\langle 88|79\rangle}(y)\left(1 - F_{Y\langle 88|79\rangle}(y)\right)}dF_{Y\langle 88|79\rangle}(y),$$

$$\int_{\mathcal{X}}\int_{\mathcal{Y}}\frac{\left(\widehat{F}_{Y_{88}|X_{88}}(y|x) - F_{Y_{88}|X_{88}}(y|x)\right)^2}{F_{Y_{88}|X_{88}}(y|x)\left(1 - F_{Y_{88}|X_{88}}(y|x)\right)}dF_{Y_{88}|X_{88}}(y|x)\,dF_{X_{88}}(x).$$

These integrals are approximated numerically by averages over 1000 points of x drawn randomly from the distribution of X_{88} and over 100 points of y corresponding to 100 quantiles of Y_{88} on a uniform grid between 0 and 1.

Table B1 to B6 report the results. Since the relative efficiency of the estimators does not vary when we consider the MSE or the weighted MSE, we will comment in detail the MSE results only. This similarity means that the relative performance of the estimators does not change much when we move from the center of the distribution to the tails.

Tables B1 and B2 provide the detailed results for the first DGP in the absence and in the presence of rounding for 100, 1,000 and 10,000 observations. When the dependent variable is continuous, the quantile regression estimator produces the most accurate estimates for both the counterfactual and conditional distributions and for all sample sizes considered. In this DGP, all estimators are correctly specified. The distribution regression estimator performs worse because it binarizes the information contained in the value of the dependent variable. Rearranging the estimates always improves the performance of the distribution regression estimator, although this improvement vanishes when the sample size increases.

The right panels of Tables B1 and B2 provide the results in the presence of discrete mass points in the wage distribution. Rounding one third of the wages to the next dollar is enough to revert the order between the MSE of the quantile and distribution regression estimators, when the number of observations is above 100 for the counterfactual distribution and when the number of observations is above 1,000 for the conditional distribution. The rounding introduces a misspecification of the linear quantile regression model while it does not affect the validity of the logit model for the conditional distribution. Thus, the linear quantile regression estimator is a better estimator when the dependent variable is continuous but a realistic amount of rounding is enough to revert the results.

One may argue that this first DGP favors the distribution regresion estimator because the error terms are logistically distributed. The results for the second DGP reported in Tables B3 and B4 show that this ingredient is not crucial. Replacing the logistic distribution with the empirical distribution of the OLS residuals in the whole population does not change qualitatively the results.

The third DGP relaxes the independence assumption between regressors and errors by introducing linear multiplicative heteroscedasticity. In the absence of rounding, the linear quantile regression estimator is still correctly specified while the logit regression is doubly misspecified (non-logistic error terms and heteroscedasticity). In this sense, the third DGP favors the quantile regression estimator. Despite of that, the results for the counterfactual distribution in Table B5 and B6 show again that the distribution regression estimator performs better than quantile regression in the presence of rounding and at least 1,000 observations. The difference is that now quantile regression is a better estimator of the conditional distribution even in the presence of rounding. In our application, the main objects of interest are counterfactual distributions, we have more than 20,000 observations in one period and more than 70,000 in the other, and the amount of rounding is at least as high as in the simulations. Therefore, even if the true DGP corresponded to this third one, we would prefer the distribution regression approach.

Since the presence of mass-points in the distribution of the dependent variable penalizes the relative performance of the quantile regression estimator, we considered applying a small random

noise to the dependent variable (also called dithering or jitering). This noise artificially restores the continuity of the distribution. The results, however, showed no improvement (even a small detorioration) in the MSE of the quantile regression estimator. On the other hand, it may be a way to restore the validity of the inference procedures.

B.4. **Conclusion.** We draw two main lessons from this section. First, the quantile regression estimator is more accurate than the distribution regression estimator for data generating processes that satisfy the assumptions of both models. Our results agree with the recent simulations of Koenker (2010). Second, introducing a realistic amount of discreteness of the dependent variable is enough to revert the results. The distribution regression approach is naturally robust to such an ubiquitous phenomenon while quantile regression is not.

Quantile and distribution regression make different parametric assumptions. It is therefore easy to find data generating processes for which one estimator dominates the other because of misspecification. This was not the goal of this simulation exercise but this is an empirical question, the response of which changes from one case to the other. In our application, for instance, we think that it is important to allow for different coefficients below and above the minimum wage. Moreover, misspecification tests reported in Rothe and Wied (2012) reject the quantile regression but not the distribution regression estimator for a similar dataset.

On the bright side, both distribution regression (independently of the link function) and quantile regression give numerically identical results in saturated models. Therefore, the choice between distribution or quantile regression becomes immaterial if we have a flexible enough specification. In our application, while the results are statistically significantly different, their economic interpretation remains extremely similar.

Appendix C. Additional Empirical Results

Table C1 contains descriptive statistics for the data sets used in CFM and in this appendix. Between 1979 and 1988 the average real wage decreased for men while it increased for women. The level of potential experience decreased because of the entry of the baby-boom generation into the labor market and because of longer education. Educational attainment increased clearly over the period. As it is well known, de-unionization was important with a 11% fall in union members for men and 5% for women. The rise of the service sector is another remarkable change that took place during this period.

Figure C1 presents robustness checks with respect to the link functions used for the distribution regression estimation. The differences between the estimates obtained with the logistic, normal, uniform (linear probability model), Cauchy and complementary log-log link functions are so modest that the lines are almost indistinguishable.

Table C2 and Figure C2 present robustness checks with respect to the wage mechanism below the minimum wage. The assumptions about the minimum wage are particularly delicate, since the mechanism that generates wages strictly below this level is not clear; it could be measurement error, non-coverage, or non-compliance with the law. To check the robustness of the results to the DFL assumptions about the minimum wage and to our semi-parametric model of the conditional distribution, we re-estimate the decomposition using censored linear quantile regression and censored distribution regression with a logit link, censoring the wage data at the level of the minimum wage. For censored quantile regression, we use Powell's (1986) censored quantile regression estimated using Chernozhukov and Hong's (2002) algorithm. For censored distribution regression, we simply censor to zero the distribution regression estimates of the conditional distributions below the minimum wage and recompute the functionals of interest. Overall, we find that the results are very similar for the quantile and distribution regressions, and they are not very sensitive to the censoring. Table C3 shows that reversing the order of the factors to: (1) labor force composition, (2) de-unionization, (4) minimum wage, and (4) wage structure, has little qualitative effect.

We present our results for female workers in Tables C4 and C5 as well as Figures C3 to C7. Table C4 reports the decomposition of the changes in various measures of wage dispersion between 1979 and 1988 estimated using logit distribution regressions. Table C5 reports the results of the same decomposition estimated using the censored distribution regression and the censored quantile regression estimators. Figures C3 to C5 refine these results by presenting estimates and 95% simultaneous confidence bands for quantile, distributional and Lorenz effects. The procedures used are exactly the same that were used to produce the tables and figures for men. Figures C6 and C7 show that this results are not sensitive to the choice of the estimator of the conditional distribution.

Most of the patterns are similar although there are a few interesting differences in the strength of the effects that explain the changes in inequality. First, de-unionization is virtually irrelevant for women. This is due to the fact that the proportion of unionized female workers has always been small. The decline in the unionization rate is smaller for females than for males in absolute value. In addition, unions do not compress the conditional female wage distribution while they reduce the conditional variance of the male wage distribution. Second, the decrease in the real value of the minimum wage explains a larger increase in wage inequality for women than for men. The reason is that the proportion of workers at or below the minimum wage is higher for female than for male workers. Thus, mechanically, women will be more affected by a decrease in the value of the minimum wage. Third, changes in individual attributes generally have a more important effect than changes in the wage structure, which is not true for men. The robustness checks using the censored distribution and quantile models as well as the different link functions for the distribution regression estimators confirm these results.

APPENDIX D. VARIANCE DECOMPOSITION INTO BETWEEN-GROUP AND WITHIN-GROUP COMPONENTS

By the law of total variance we can decompose the variance as

$$\text{Var}\left(Y\right) = \text{Var}\left[E\left(Y|X\right)\right] + E\left[\text{Var}\left(Y|X\right)\right],$$

where the first term is the between-group variance and the second term the within-group variance. The regression models for conditional distributions considered in CFM naturally lead to models for the conditional mean and variance of Y given X. By definition,

$$E\left(Y|X = x\right) = \int y \, dF_{Y|X}\left(y|x\right) \text{ and}$$

$$\text{Var}\left(Y|X = x\right) = \int \left(y - E\left(Y|X = x\right)\right)^2 dF_{Y|X}\left(y|x\right).$$

For conditional quantile models, the direct expressions are

$$E\left(Y|X = x\right) = \int_0^1 Q_{Y|X}\left(u|x\right) du \text{ and}$$

$$\text{Var}\left(Y|X = x\right) = \int_0^1 \left(Q_{Y|X}\left(u|x\right) - E\left(Y|X = x\right)\right)^2 du.$$

For linear quantile regression models, this simplifies to

$$E\left(Y|X = x\right) = x' E\left[\beta\left(u\right)\right] \text{ and}$$
$$\text{Var}\left(Y|X = x\right) = x' \text{Var}\left[\beta\left(u\right)\right] x,$$

such that the variance decomposition take the following simple form (equation (6.7) in CFM):

$$\text{Var}[Y] = E[\beta(U)]' \text{Var}[X] E[\beta(U)] + \text{trace}\{E[XX'] \text{Var}[\beta(U)]\}.$$

Both the between and the within components are functions of the conditional distribution $F_{Y|X}$ (that determines $E\left(Y|X = x\right)$ and $Var\left(Y|X = x\right)$) and the covariate distribution F_X. Models for conditional distributions allow changing only F_X while keeping $F_{Y|X}$ fixed. In general, the between and within components of the variance of the counterfactual outcome $Y\left\langle j, k\right\rangle$ are

$$\text{Var}\left[E\left(Y\left\langle j, k\right\rangle |X\right)\right] = \int \left[E\left(Y_j|X_j = x\right) - E\left(Y\left\langle j, k\right\rangle\right)\right]^2 dF_{X_k}\left(x\right) \text{ and}$$

$$E\left[\text{Var}\left(Y\left\langle j, k\right\rangle |X\right)\right] = \int Var\left(Y_j|X_j = x\right) dF_{X_k}\left(x\right).$$

In Section 6 of CFM, the whole composition effect (i.e. the effect of changing the distribution of both union status and other individual characteristics) is defined as the difference between

$F_{Y\langle(1,0)|(1,1)\rangle}$ and $F_{Y\langle(1,0)|(0,0)\rangle}$. Therefore, the composition effects on between-group and within-group inequality are

$$\mathrm{Var}\left[E\left(Y\langle(1,0)|(1,1)\rangle|X\right)\right] - \mathrm{Var}\left[E\left(Y\langle(1,0)|(0,0)\rangle|X\right)\right] \ \text{and}$$

$$E\left[\mathrm{Var}\left(Y\langle(1,0)|(1,1)\rangle|X\right)\right] - E\left[\mathrm{Var}\left(Y\langle(1,0)|(0,0)\rangle|X\right)\right].$$

When we use the logit distribution regression model to estimate the conditional distribution $F_{Y_{(1,0)}|X_1}$, we find that between-group inequality increased from 0.147 to 0.163 as a consequence of the composition changes, which represents an increase of 10.6%. Similarly, within-group inequality increased from 0.125 to 0.136, which represents an increase of 9.1%. Note that the dependent variable is the hourly log-wage.

When we use the censored quantile regression model to estimate the conditional distribution censoring wages below the minimum wage, we obtain increases in between-group and within-group inequality from 0.135 to 0.145 (7.5%) and from 0.120 to 0.131 (9.1%), respectively. If we do not censor wages below the minimum wage, then we can use directly the expressions in equation (6.7). In this case we obtain increases in for between-group and within-group inequality from 0.201 to 0.217 (7.9%) and from 0.152 to 0.164 (8.4%), respectively.

REFERENCES

[1] Beran, R. (1984): "Bootstrap Methods in Statistics," *Jahresber. Deutsch. Math. Verein.* 86, pp. 14–30.

[2] Chernozhukov, V., and I. Fernández-Val (2005): "Subsampling Inference on Quantile Regression Processes," *Sankhyā* 67, pp. 253–276.

[3] Chernozhukov, V., I. Fernandez-Val and A. Galichon (2010): "Quantile and Probability Curves without Crossing," *Econometrica*.

[4] Chernozhukov, V., I. Fernández-Val and B. Melly (2012): "Inference on Counterfactual Distributions," unpublished manuscript, MIT.

[5] Chernozhukov, V., and H. Hong (2002): "Three-step censored quantile regression and extramarital affairs," *Journal of the American Statistical Association* 97, pp. 872–882.

[6] Davydov, Yu. A., M. A. Lifshits and N. V. Smorodina (1998): Local properties of distributions of stochastic functionals. Translated from the 1995 Russian original by V. E. Nazaĭkinskiĭ and M. A. Shishkova. Translations of Mathematical Monographs, 173. American Mathematical Society, Providence, RI.

[7] Koenker, R. (2010): "On Distributional vs. Quantile Regressions," unpublished manuscript, University of Illinois at Urbana-Champaign.

[8] Koenker R. and G. Bassett (1978): "Regression Quantiles," *Econometrica* 46, pp. 33–50.

[9] Powell, J. L. (1986): "Censored Regression Quantiles," *Journal of Econometrics* 32, pp. 143-155.

[10] Rothe, C., and D. Wied (2012): "Specification Testing in a Class of Conditional Distributional Models," *Journal of the American Statistical Association*, forthcoming.

[11] van der Vaart, A., and J. Wellner (1996): *Weak convergence and empirical processes: with applications to statistics*, New York: Springer.
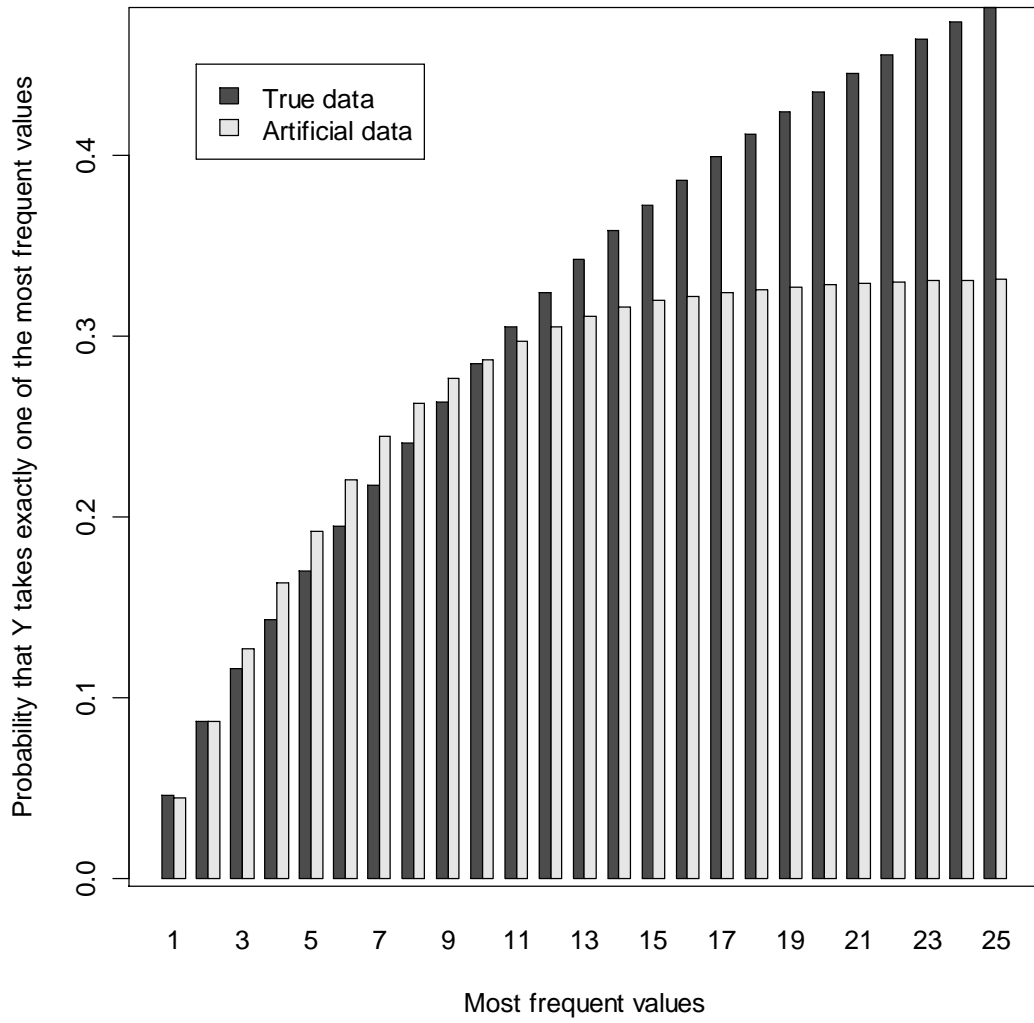
Figure B1: Amount of discreteness in the real and simulated data
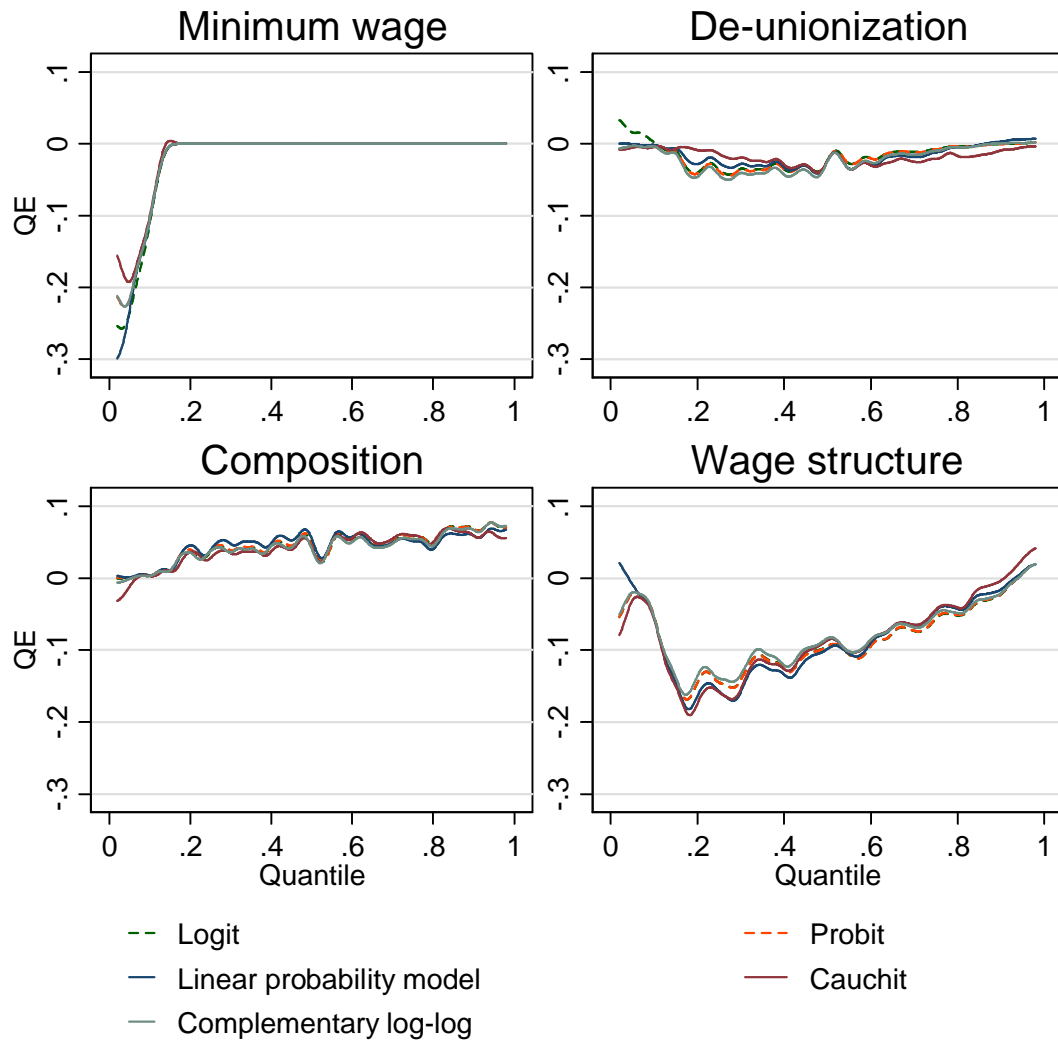
Figure C1. Comparison of the distribution regression estimates of the quantile effects based on five different link functions: logistic, normal, uniform, Cauchy and complementary log-log.
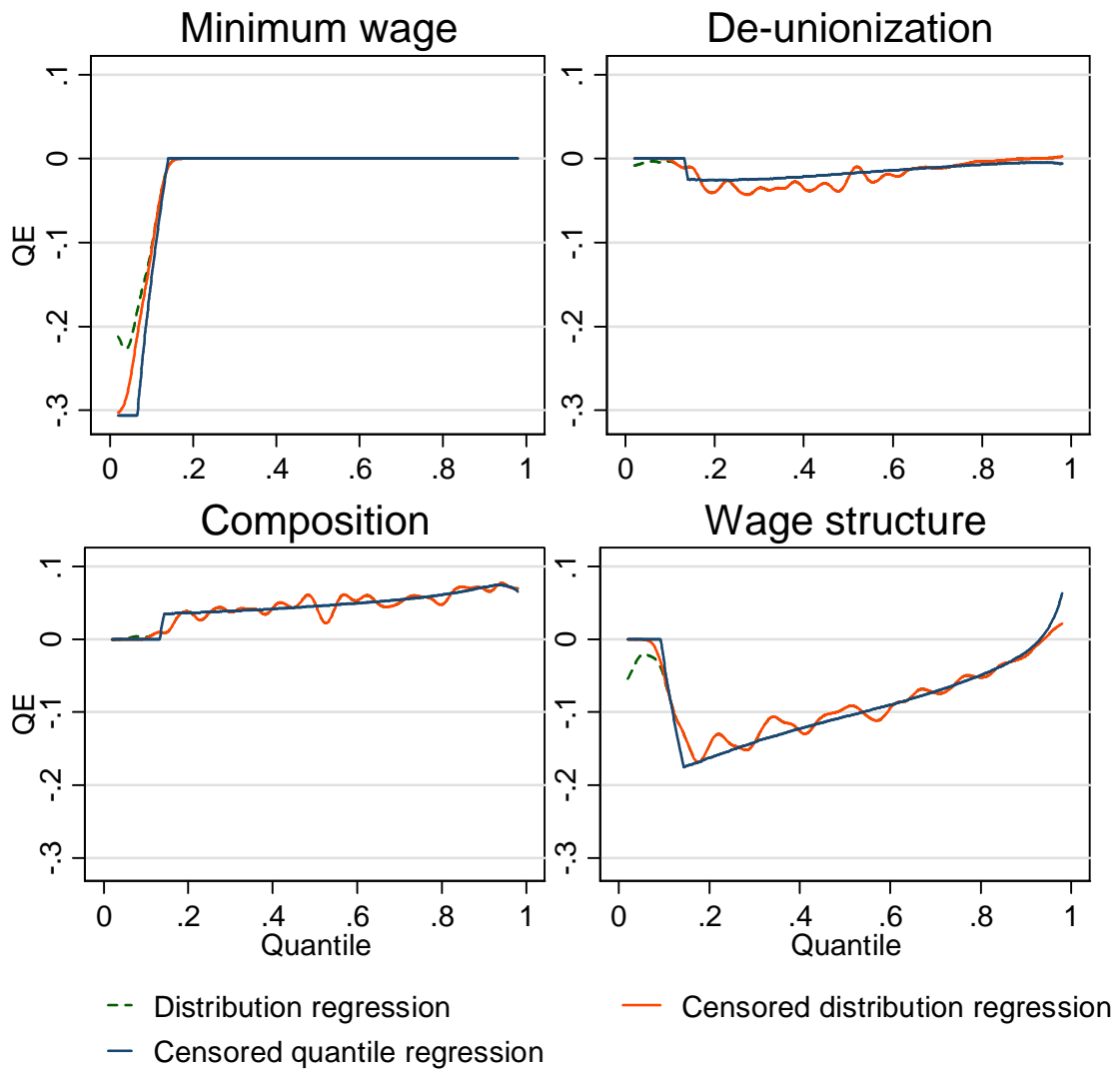
Figure C2. Comparison of the distribution regression, censored distribution regression and censored quantile regression estimates of the quantile effects.
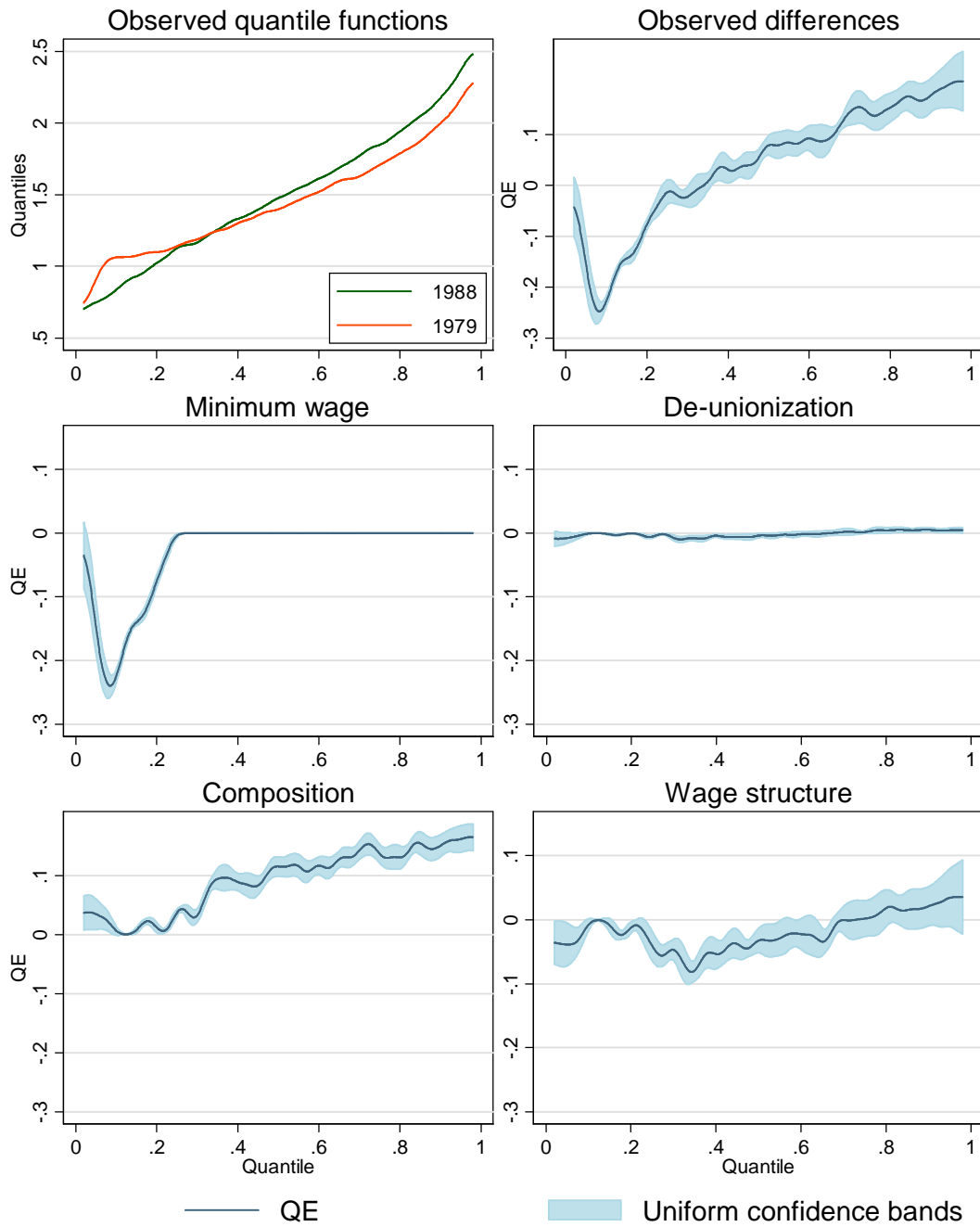
Figure C3. Observed quantile functions, observed differences between the quantile functions and their decomposition into four quantile effects. The 95% simultaneous confidence bands were obtained by empirical bootstrap with 100 repetitions. Results for women.
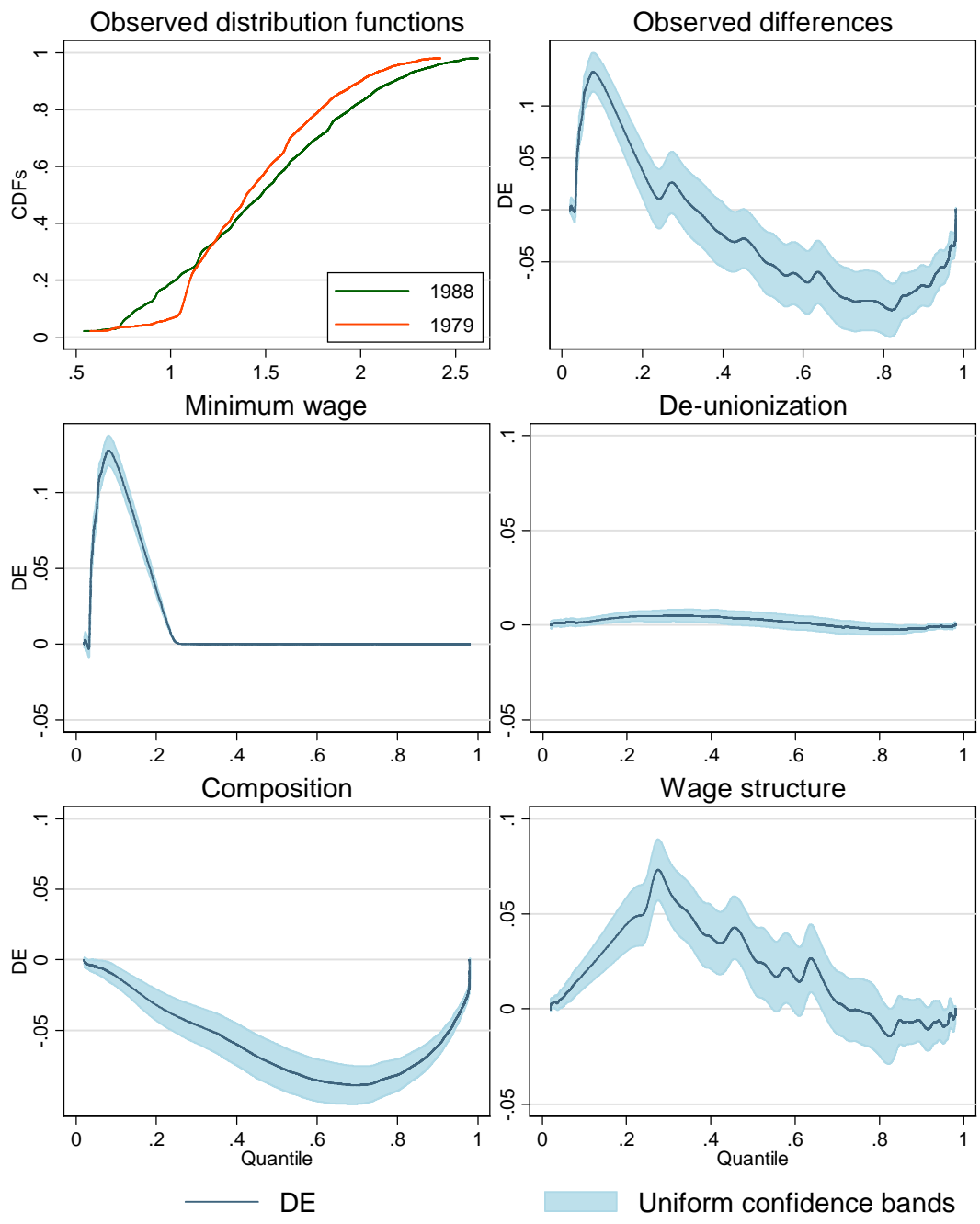
Figure C4: Observed distribution functions, observed differences between the distribution functions and their decomposition into four distribution effects. The 95% simultaneous confidence bands were obtained by empirical bootstrap with 100 repetitions. Results for women.
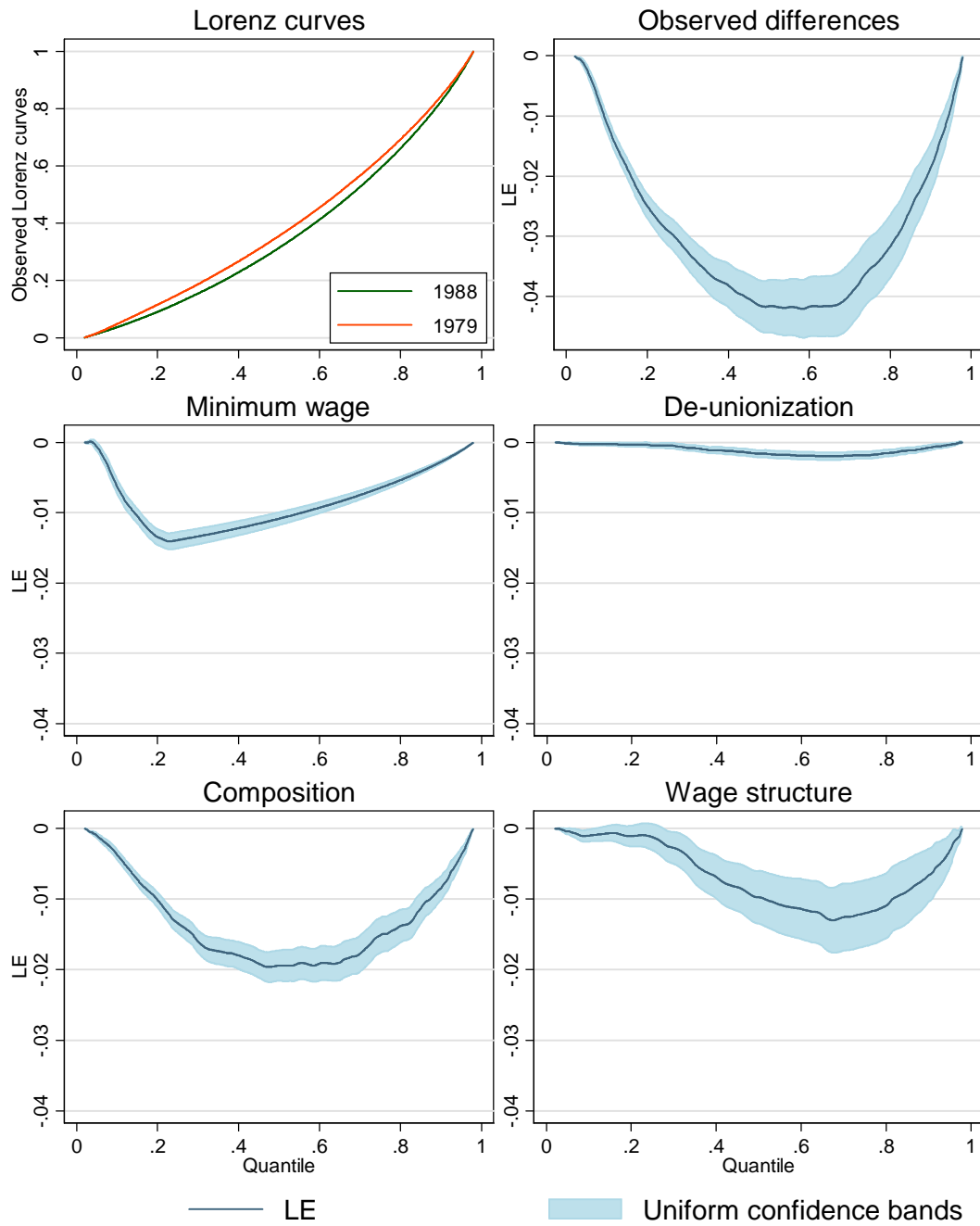
Figure C5. Observed Lorenz curves, observed differences between the Lorenz curves and their decomposition into four Lorenz effects. The 95% simultaneous confidence bands were obtained by empirical bootstrap with 100 repetitions. Results for women.
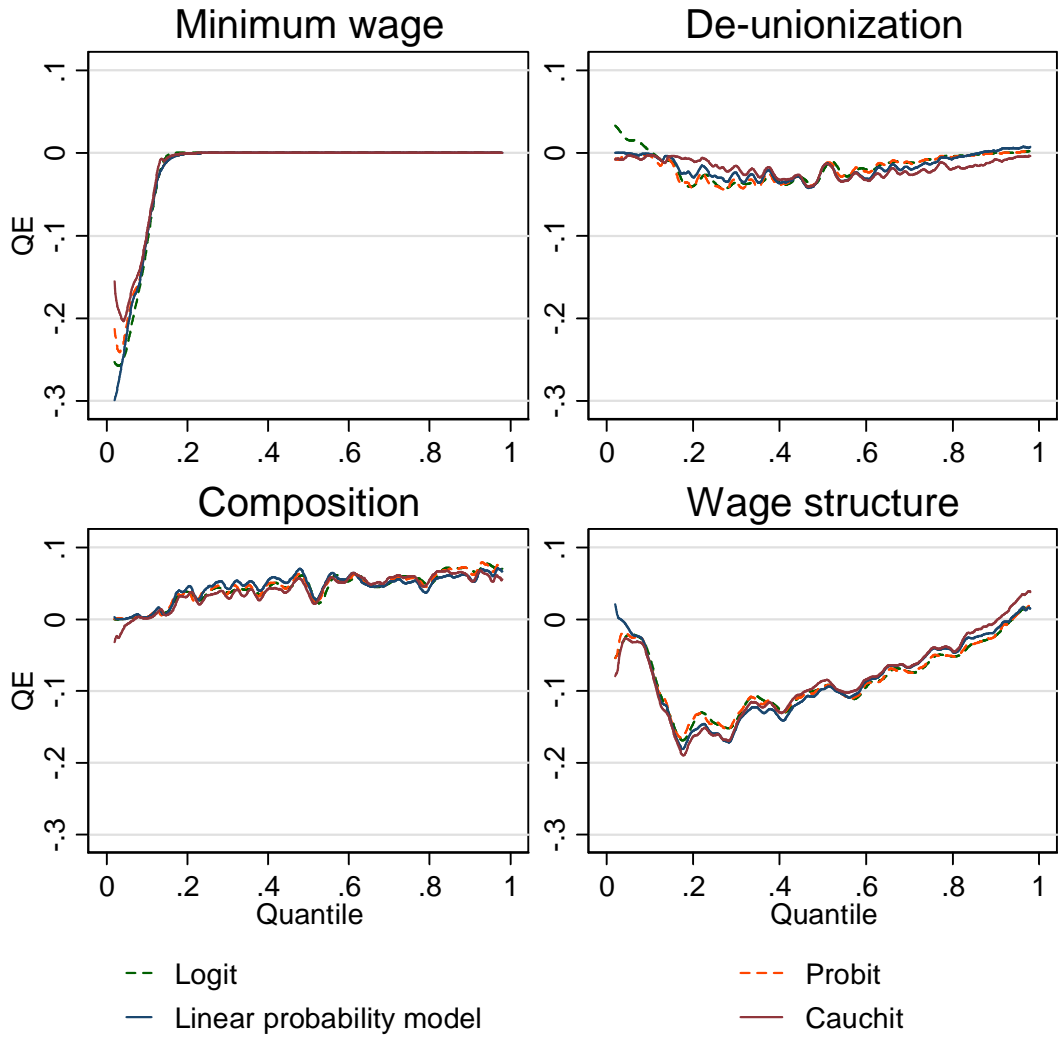
Figure C6. Comparison of the distribution regression estimates of the quantile effects based on five different link functions: logistic, normal, uniform, Cauchy and complementary log-log. Results for women.
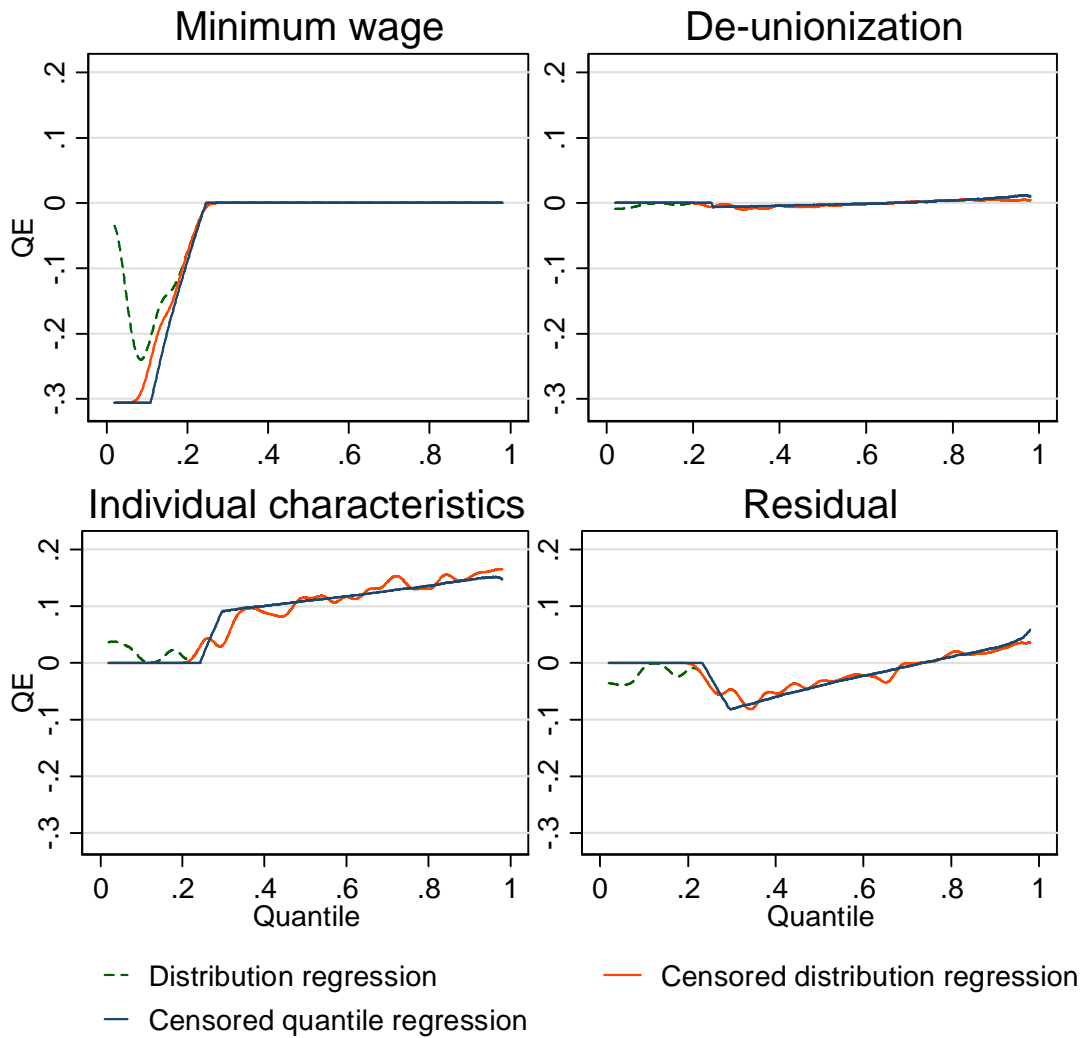
Figure C7. Comparison of the distribution regression, censored distribution regression and censored quantile regression estimates of the quantile effects. Results for women.

Table B1: Square root MSE of the distribution and quantile estimators for the DGP 1

Estimand	# of obs.	No rounding			Rounding with Pr=1/3		
		$\hat{F}_Y^{DR}$	$\hat{F}_Y^{*}$	$\hat{F}_Y^{QR}$	$\hat{F}_Y^{DR}$	$\hat{F}_Y^{*}$	$\hat{F}_Y^{QR}$
Counterfactual distribution	100	5.09	4.73	3.96	5.53	5.16	4.50
	1000	1.42	1.41	1.19	1.54	1.54	1.86
	10000	0.46	0.46	0.39	0.50	0.50	1.40
Conditional distribution	100	18.76	16.11	12.41	20.35	17.56	13.68
	1000	5.11	4.94	3.86	5.55	5.38	4.66
	10000	1.57	1.57	1.23	1.72	1.71	2.44

Note: All numbers are multiplied by 1000

Table B2: Square root Anderson-Darling weighted MSE of the distribution and quantile estimators for the DGP 1

Estimand	# of obs.	No rounding			Rounding with Pr=1/3		
		$\hat{F}_Y^{DR}$	$\hat{F}_Y^{*}$	$\hat{F}_Y^{QR}$	$\hat{F}_Y^{DR}$	$\hat{F}_Y^{*}$	$\hat{F}_Y^{QR}$
Counterfactual distribution	100	14.80	11.55	9.64	16.35	12.60	10.95
	1000	3.48	3.44	2.86	3.78	3.75	4.36
	10000	1.12	1.12	0.93	1.22	1.22	3.22
Conditional distribution	100	54.51	34.38	30.39	59.87	42.03	33.58
	1000	12.85	11.85	9.46	14.05	12.95	11.28
	10000	3.78	3.76	3.03	4.14	4.12	5.64

Note: All numbers are multiplied by 1000

Table B3: Square root MSE of the distribution and quantile estimators for the DGP 2

Estimand	# of obs.	No rounding			Rounding with Pr=1/3		
		$\hat{F}_Y^{DR}$	$\hat{F}_Y^{*}$	$\hat{F}_Y^{QR}$	$\hat{F}_Y^{DR}$	$\hat{F}_Y^{*}$	$\hat{F}_Y^{QR}$
Counterfactual distribution	100	5.04	4.69	3.99	5.46	5.09	4.5
	1000	1.44	1.43	1.19	1.56	1.55	1.8
	10000	0.46	0.46	0.39	0.5	0.5	1.33
Conditional distribution	100	18.62	15.9	12.29	20.24	17.38	13.54
	1000	5.11	4.93	3.87	5.56	5.38	4.62
	10000	1.62	1.61	1.23	1.77	1.76	2.32

Note: All numbers are multiplied by 1000

Table B4: Square root Anderson-Darling weighted MSE of the distribution and quantile estimators for the DGP 2

Estimand	# of obs.	No rounding			Rounding with Pr=1/3		
		$\hat{F}_Y^{DR}$	$\hat{F}_Y^{*}$	$\hat{F}_Y^{QR}$	$\hat{F}_Y^{DR}$	$\hat{F}_Y^{*}$	$\hat{F}_Y^{QR}$
Counterfactual distribution	100	14.51	11.45	9.66	16.02	12.48	10.95
	1000	3.51	3.47	2.87	3.83	3.79	4.28
	10000	1.12	1.12	0.94	1.23	1.23	3.11
Conditional distribution	100	54.05	38.00	30.20	59.67	41.73	33.37
	1000	12.84	11.87	9.49	14.11	13.01	11.23
	10000	3.88	3.87	3.04	4.28	4.26	5.43

Note: All numbers are multiplied by 1000

Table B5: Square root MSE of the distribution and quantile estimators for the DGP 3

Estimand	# of obs.	No rounding			Rounding with Pr=1/3		
		$\hat{F}_Y^{DR}$	$\hat{F}_Y^{*}$	$\hat{F}_Y^{QR}$	$\hat{F}_Y^{DR}$	$\hat{F}_Y^{*}$	$\hat{F}_Y^{QR}$
Counterfactual distribution	100	4.98	4.63	3.98	5.43	5.06	4.52
	1000	1.43	1.42	1.18	1.55	1.54	1.86
	10000	0.5	0.5	0.39	0.55	0.55	1.42
Conditional distribution	100	18.94	16.16	12.27	20.62	17.68	13.56
	1000	5.75	5.58	3.9	6.19	6.03	4.74
	10000	3.02	3.02	1.29	3.17	3.17	2.54

Note: All numbers are multiplied by 1000

Table B6: Square root Anderson-Darling weighted MSE of the distribution and quantile estimators for the DGP 3

Estimand	# of obs.	No rounding			Rounding with Pr=1/3		
		$\hat{F}_Y^{DR}$	$\hat{F}_Y^{*}$	$\hat{F}_Y^{QR}$	$\hat{F}_Y^{DR}$	$\hat{F}_Y^{*}$	$\hat{F}_Y^{QR}$
Counterfactual distribution	100	14.35	11.27	9.63	15.77	12.33	10.96
	1000	3.46	3.42	2.85	3.76	3.73	4.40
	10000	1.19	1.19	0.93	1.29	1.29	3.30
Conditional distribution	100	54.95	39.01	30.54	60.41	42.80	33.76
	1000	14.64	13.74	9.67	15.76	14.81	11.58
	10000	7.66	7.65	3.31	7.98	7.97	6.00

Note: All numbers are multiplied by 1000

Table C1: Summary statistics

	Males		Females	
Variable	1979	1988	1979	1988
Number of obs.	21,483	74,661	16,911	70,089
Wage	7.24 (3.71)	7.01 (4.32)	4.75 (2.39)	5.05 (2.99)
Education	12.44 (2.98)	12.95 (2.80)	12.47 (2.48)	13.02 (2.44)
Experience	18.69 (13.55)	18.02 (12.25)	18.28 (13.66)	17.81 (12.39)
Union	31.77 (46.56)	20.58 (40.43)	17.21 (37.75)	12.65 (33.24)
Primary sector	4.59 (20.93)	3.81 (19.14)	1.20 (10.89)	1.06 (10.22)
Secondary sector	39.99 (48.99)	34.90 (47.67)	19.06 (39.28)	15.20 (35.90)
Tertiary sector	55.42 (49.71)	61.29 (48.71)	79.74 (40.19)	83.74 (36.90)
Part-time	6.93 (25.39)	10.55 (30.73)	22.20 (41.56)	25.30 (43.47)
Non-white	9.79 (29.72)	11.30 (31.66)	13.09 (33.73)	13.68 (34.37)
SMSA	59.17 (49.15)	73.09 (44.35)	60.84 (48.81)	73.72 (44.02)

Notes: Mean of selected variables. The standard deviations are reported in parenthesis. When the variable is binary, the results are reported in percentage.

Table C2: Decomposing changes in measures of wage dispersion: men, censored models

Statistic	Total change	Minimum wage	Unions	Composition	Wage structure
		Effect of:			
CDR:					
Standard	8.2 (0.3)	3.3 (0.0)	0.6 (0.0)	1.9 (0.2)	2.4 (0.2)
Deviation		40.7 (1.4)	7.9 (0.5)	22.5 (1.8)	28.9 (2.4)
90-10	21.5 (1.0)	11.2 (0.1)	0.0 (0.0)	9.2 (0.8)	1.1 (1.3)
		52.1 (2.4)	0.0 (0.1)	42.6 (4.4)	5.3 (5.9)
50-10	11.3 (1.4)	11.2 (0.1)	-2.0 (1.0)	5.1 (0.4)	-3.1 (1.1)
		99.6 (14.1)	-17.9 (11.2)	45.5 (8.3)	-27.2 (14.0)
90-50	10.2 (1.2)	0.0 (0.0)	2.0 (1.0)	4.0 (0.8)	4.2 (1.1)
		0.0 (0.0)	19.7 (8.4)	39.3 (8.8)	41.0 (9.8)
75-25	15.4 (1.1)	0.0 (0.0)	4.1 (1.0)	0.3 (1.3)	11.1 (1.2)
		0.0 (0.0)	26.5 (6.2)	1.7 (8.6)	71.8 (8.7)
95-5	36.4 (2.1)	26.4 (0.7)	0.0 (0.6)	8.5 (1.1)	1.4 (1.5)
		72.7 (3.8)	0.0 (1.5)	23.4 (2.7)	3.9 (4.0)
Gini	4.2 (0.1)	1.6 (0.0)	0.4 (0.0)	0.3 (0.1)	1.8 (0.1)
coefficient		37.9 (1.1)	10.7 (0.5)	7.1 (1.6)	44.2 (1.6)
CQR:					
Standard	9.0 (0.3)	2.8 (0.2)	4.1 (0.0)	0.3 (0.0)	1.8 (0.1)
Deviation		45.3 (1.5)	3.2 (0.5)	20.0 (1.6)	31.5 (2.2)
90-10	22.3 (1.1)	1.4 (1.1)	14.2 (0.4)	-0.5 (0.1)	7.2 (0.4)
		63.6 (3.4)	-2.2 (0.6)	32.3 (2.8)	6.4 (5.1)
50-10	9.5 (0.9)	14.2 (0.4)	-1.8 (0.1)	4.6 (0.4)	-7.4 (0.9)
		148.7 (16.7)	-18.7 (3.0)	47.9 (9.0)	-78.0 (21.6)
90-50	12.7 (0.7)	0.0 (0.0)	1.3 (0.1)	2.6 (0.3)	8.8 (0.5)
		0.0 (0.0)	10.1 (1.0)	20.6 (2.4)	69.4 (2.5)
75-25	12.7 (0.6)	0.0 (0.0)	1.6 (0.1)	2.0 (0.4)	9.1 (0.5)
		0.0 (0.0)	12.9 (1.2)	15.5 (3.0)	71.5 (3.1)
95-5	39.2 (0.8)	30.6 (0.0)	-0.5 (0.1)	7.4 (0.5)	1.6 (0.8)
		78.1 (1.8)	-1.2 (0.3)	18.9 (1.2)	4.2 (2.1)
Gini	4.5 (0.1)	1.9 (0.0)	0.3 (0.0)	0.3 (0.1)	2.1 (0.1)
coefficient		42.2 (1.1)	5.9 (0.4)	6.1 (1.4)	45.9 (1.4)

Notes: The censored logit distribution regression and the censored linear quantile regression estimators have been applied. All numbers are in %. Bootstrapped standard errors with 100 repetitions are given in parenthesis. The second line in each cell indicates the percentage of total variation.

Table C3: Reversing the order of the decomposition

Statistic	Total change	Effect of: Individual attributes	Unions	Minimum wage	Coefficients
Men:					
Standard Deviation	8.0 (0.3)	0.9 (0.2)	1.5 (0.1)	2.9 (0.2)	2.7 (0.3)
		11.4 (2.8)	19.2 (1.0)	36.3 (2.8)	33.1 (2.4)
90-10	21.5 (1.0)	0.4 (1.3)	8.8 (1.2)	11.2 (0.9)	1.1 (1.3)
		1.8 (5.8)	40.7 (5.6)	52.1 (4.8)	5.3 (5.9)
50-10	11.3 (1.4)	2.5 (1.5)	0.7 (1.2)	11.2 (0.5)	-1.4 (0.7)
		21.8 (12.7)	5.8 (11.3)	99.6 (15.5)	-3.1 (1.1)
90-50	10.2 (1.2)	-2.1 (1.2)	8.1 (0.8)	0.0 (0.9)	4.2 (1.1)
		-20.1 (12.9)	79.1 (0.7)	0.0 (9.5)	41.0 (9.8)
75-25	15.4 (1.1)	6.4 (1.2)	-2.1 (1.3)	0.0 (1.0)	11.1 (1.2)
		41.6 (6.8)	-13.4 (9.2)	0.0 (7.4)	71.8 (8.7)
95-5	33.0 (2.1)	2.6 (1.7)	5.9 (1.1)	23.0 (1.0)	1.4 (1.5)
		7.9 (4.8)	17.9 (3.4)	69.9 (5.5)	4.3 (4.4)
Gini coefficient	4.1 (0.1)	-0.3 (0.1)	1.0 (0.0)	1.4 (0.1)	2.0 (0.1)
		-6.8 (2.8)	24.1 (1.1)	33.3 (2.5)	49.4 (1.8)
Women:					
Standard Deviation	10.9 (0.4)	4.5 (0.2)	0.0 (0.0)	4.4 (0.3)	2.1 (0.3)
		41.1 (2.4)	-0.2 (0.2)	40.0 (2.4)	19.1 (2.5)
90-10	39.8 (1.4)	11.2 (0.7)	0.0 (0.4)	27.2 (0.2)	1.3 (1.1)
		28.2 (1.6)	0.0 (1.0)	68.4 (2.3)	3.4 (2.6)
50-10	33.0 (0.7)	7.9 (0.8)	-0.8 (0.5)	22.7 (0.5)	-1.4 (0.7)
		24.1 (2.6)	-2.4 (1.7)	82.6 (2.2)	-4.3 (2.4)
90-50	6.8 (1.4)	3.3 (0.8)	0.8 (0.5)	0.0 (0.5)	2.8 (1.4)
		47.9 (13.4)	11.7 (7.2)	0.0 (6.5)	40.3 (9.9)
75-25	12.8 (0.9)	2.8 (0.7)	0.0 (0.5)	5.5 (0.3)	4.5 (0.8)
		22.0 (5.6)	0.0 (3.9)	43.0 (3.5)	35.0 (4.5)
95-5	38.8 (1.9)	17.4 (0.9)	0.0 (0.3)	16.5 (2.0)	5.0 (2.1)
		44.8 (3.0)	0.0 (0.9)	42.4 (5.1)	12.8 (5.1)
Gini coefficient	4.0 (0.1)	0.6 (0.1)	0.0 (0.0)	2.5 (0.1)	0.9 (0.1)
		14.7 (2.4)	1.2 (0.3)	61.1 (2.7)	23.0 (2.2)

Notes: The logit distribution regression model has been applied. All numbers are in %. Bootstrapped standard errors with 100 repetitions are given in parenthesis. The second line in each cell indicates the percentage of total variation.

Table C4: Decomposing changes in measures of wage dispersion: women, distribution regression

Statistic	Total change	Effect of:			
		Minimum wage	Unions	Individual attributes	Coefficients
Standard	10.9 (0.4)	3.8 (0.1)	0.3 (0.0)	4.7 (0.2)	2.1 (0.3)
Deviation		34.9 (1.5)	3.2 (0.4)	42.8 (1.8)	19.1 (2.5)
90-10	39.8 (1.4)	23.0 (0.2)	0.9 (0.5)	14.5 (0.7)	1.3 (1.1)
		57.9 (1.9)	2.3 (1.2)	36.4 (1.7)	3.4 (2.6)
50-10	33.0 (0.7)	23.0 (0.2)	0.0 (0.1)	11.3 (0.4)	-1.4 (0.7)
		69.9 (1.6)	0.0 (0.4)	34.4 (1.3)	-4.3 (2.4)
90-50	6.8 (1.4)	0.0 (0.0)	0.9 (0.5)	3.1 (0.8)	2.8 (1.4)
		0.0 (0.0)	13.6 (7.2)	46.0 (11.3)	40.3 (9.9)
75-25	12.8 (0.9)	0.0 (0.0)	0.0 (0.5)	8.3 (0.2)	4.5 (0.8)
		0.0 (0.0)	0.0 (3.9)	65.1 (5.0)	35.0 (4.5)
95-5	38.8 (1.9)	16.8 (0.5)	0.7 (0.7)	16.4 (2.0)	5.0 (2.1)
		43.2 (2.2)	1.9 (1.9)	42.1 (5.0)	12.8 (5.1)
Gini	4.0 (0.1)	2.0 (0.1)	0.1 (0.0)	1.0 (0.1)	0.9 (0.1)
coefficient		49.0 (1.8)	3.5 (0.4)	24.5 (1.4)	23.0 (2.2)

Notes: The logit distribution regression model has been applied. All numbers are in %. Bootstrapped standard errors with 100 repetitions are given in parenthesis. The second line in each cell indicates the percentage of total variation.

Table C5: Decomposing changes in measures of wage dispersion: women, censored models

Statistic	Total change	Minimum wage	Unions	Individual attributes	Coefficients
CDR:					
Standard	12.7 (0.4)	5.6 (0.1)	0.3 (0.0)	1.9 (0.2)	2.4 (0.2)
Deviation		44.1 (1.5)	2.2 (0.4)	39.9 (1.8)	13.8 (2.5)
90-10	43.2 (1.4)	26.4 (0.2)	0.9 (0.5)	14.5 (0.7)	13.0 (2.6)
		61.2 (1.9)	2.2 (1.2)	33.5 (1.7)	3.1 (2.6)
50-10	36.4 (0.7)	26.4 (0.2)	0.0 (0.1)	11.3 (0.4)	-1.4 (0.7)
		72.7 (1.6)	0.0 (0.4)	31.2 (1.3)	-3.9 (2.4)
90-50	6.8 (1.4)	0.0 (0.0)	0.9 (0.5)	3.1 (0.8)	2.8 (1.4)
		0.0 (0.0)	13.6 (7.2)	46.0 (11.3)	40.3 (9.9)
75-25	12.8 (0.9)	0.0 (0.0)	0.0 (0.5)	8.3 (0.2)	4.5 (0.8)
		0.0 (0.0)	0.0 (3.9)	65.1 (5.0)	35.0 (4.5)
95-5	52.7 (1.9)	30.6 (0.5)	0.7 (0.7)	16.7 (2.0)	4.7 (2.1)
		58.1 (2.2)	1.4 (1.9)	31.6 (5.0)	8.8 (5.1)
Gini	4.9 (0.1)	2.9 (0.1)	0.1 (0.0)	1.3 (0.1)	0.6 (0.1)
coefficient		59.2 (1.8)	1.9 (0.4)	26.1 (1.4)	12.8 (2.2)
CQR:					
Standard	13.1 (0.4)	6.2 (0.0)	0.3 (0.1)	4.5 (0.3)	2.0 (0.3)
Deviation		47.6 (1.6)	2.6 (0.4)	34.8 (1.5)	15.0 (1.9)
90-10	48.8 (1.2)	30.6 (0.0)	0.8 (0.2)	14.7 (0.8)	2.7 (0.8)
		62.8 (1.5)	1.6 (0.3)	30.1 (1.3)	5.5 (1.6)
50-10	37.2 (0.7)	30.6 (0.0)	-0.3 (0.1)	10.9 (0.8)	-4.1 (0.5)
		82.3 (1.6)	-0.7 (0.3)	29.4 (1.7)	-10.9 (1.4)
90-50	11.5 (0.9)	0.0 (0.0)	1.1 (0.1)	3.7 (0.5)	6.7 (0.8)
		0.0 (0.0)	9.1 (1.1)	32.3 (3.3)	58.5 (3.5)
75-25	15.3 (0.9)	0.0 (0.0)	0.8 (0.1)	11.8 (0.8)	2.6 (0.9)
		0.0 (0.0)	5.6 (0.7)	77.6 (5.2)	16.9 (5.1)
95-5	50.8 (1.4)	30.6 (0.0)	1.1 (0.2)	15.1 (0.8)	4.0 (1.0)
		60.3 (1.6)	2.1 (0.4)	29.7 (1.2)	7.9 (1.8)
Gini	5.1 (0.1)	3.2 (0.0)	0.1 (0.0)	1.1 (0.1)	0.7 (0.1)
coefficient		62.5 (1.5)	2.1 (0.3)	21.8 (1.2)	13.6 (1.5)

Notes: All numbers are in %. Bootstrapped standard errors with 100 repetitions are given in parenthesis. The second line in each cell indicates the percentage of total variation. The censored distribution regression estimator with logistic link and the censored linear quantile regression estimator have been applied.