Chernozhukov, Victor; Fernández-Val, Iván; Melly, Blaise

Working Paper

Inference on counterfactual distributions

cemmap working paper, No. CWP05/12

Provided in Cooperation with:
The Institute for Fiscal Studies (IFS), London

Suggested Citation: Chernozhukov, Victor; Fernández-Val, Iván; Melly, Blaise (2012) : Inference on counterfactual distributions, cemmap working paper, No. CWP05/12, Centre for Microdata Methods and Practice (cemmap), London, https://doi.org/10.1920/wp.cem.2012.0512

This Version is available at:
https://hdl.handle.net/10419/64695

cemmap
centre for microdata methods and practice

Inference on counterfactual distributions

Victor Chernozhukov
Iván Fernández-Val
Blaise Melly

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP05/12

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL
An ESRC Research Centre

INFERENCE ON COUNTERFACTUAL DISTRIBUTIONS

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

Abstract. We develop inference procedures for policy analysis based on regression methods. We consider policy interventions that correspond to either changes in the distribution of covariates, or changes in the conditional distribution of the outcome given covariates, or both. Under either of these policy 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 policy effects, including the effects on the marginal distribution function, quantile function, and other related functionals. We use these confidence sets to test functional hypotheses such as no-effect, positive effect, or stochastic dominance. Our theory applies to general policy interventions and covers the main regression methods including classical, quantile, duration, and distribution regressions. We illustrate the results with an empirical application on wage decompositions using data for the United States. Of independent interest is the use of distribution regression as a tool for modeling the entire conditional distribution, encompassing duration/transformation regression, and representing an alternative to quantile regression.

Key Words: Policy effects, counterfactual distribution, quantile regression, distribution regression, duration/transformation regression

Date: February 8, 2012. First ArXiv version: April 6, 2009. This paper replaces the earlier independent projects started in 2005 "Inference on Counterfactual Distributions Using Conditional Quantile Models," by Chernozhukov and Fernández-Val, and "Estimation of Counterfactual Distributions Using Quantile Regression," by Melly. We would like to thank the editor, three anonymous referees, Alberto Abadie, Isaiah Andrews, Josh Angrist, Manuel Arellano, David Autor, Alexandre Belloni, Arun Chandrasekhar, Denis Chetverikov, Flavio Cunha, Brigham Frandsen, Jerry Hausman, Jim Heckman, Michael Jansson, Kengo Kato, Roger Koenker, Joonhwan Lee, Ye Luo, Pierre-Andre Maugis, Justin McCrary, and seminar participants at Banff International Research Station Conference on Semiparameteric and Nonparametric Methods in Econometrics, Berkeley, Boston University, CEMFI, Columbia, Harvard/MIT, Michigan, MIT, Ohio State, St. Gallen, and 2008 Winter Econometric Society Meetings for very useful comments that helped improve the paper. Companion software developed by the authors (counterfactual package for Stata) is available from Blaise Melly. We gratefully acknowledge research support from the NSF.

1. Introduction

Policy analysis in economics aims to predict the effect of a potential policy intervention or a counterfactual change in economic conditions on some outcome variable of interest (e.g., Stock, 1989, Juhn, Murphy and Pierce, 1993, Abbring and Heckman, 2007). For example, we might be interested in what the wage distribution would be in 2000 if workers have the same characteristics as in 1990. Or we might be interested in what the distribution of wages for female workers would be in the absence of gender discrimination in the labor market. More generally, we can often think of a policy intervention either as 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, or both. Policy analysis consists of estimating the effect of such policy interventions on the marginal distribution of Y.

The main objective and contribution of this paper is to develop inference procedures for policy analysis based on regression methods. Starting from regression estimates of the conditional distribution of the outcome given covariates and nonparametric estimates of the covariate distribution, we obtain uniformly consistent and asymptotically Gaussian estimates for "policy functionals" – namely, functionals of the marginal distribution of the outcome before and after the policy intervention. Examples of these policy functionals include distribution functions, quantile functions, quantile policy effects, distribution policy effects, Lorenz curves, and Gini coefficients. We then construct confidence sets around these estimates 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 policy 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 principal regression methods for estimating conditional distributions most commonly used in empirical work, including classical, quantile, duration/transformation, and distribution regressions. We consider simple interventions consisting of marginal changes in the values of a given covariate, as well as more elaborate policies consisting of general changes in the covariate distribution or in the conditional distribution of the outcome given covariates. Moreover, 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 policy 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 policy 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 policy functionals. The key to all these results is the Hadamard differentiability of the policy functionals with respect to the first stage function-valued parameters, which we establish in the paper. Given this key ingredient, all of the results above follow from the functional delta method. An 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 our leading examples. Specifically, we verify the high-level conditions – functional central limit theorem and validity of bootstrap – 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 processes of distribution regression. We also prove the validity of the exchangeable bootstrap for the empirical coefficient processes 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 policy analysis based on regression methods. Stock (1989) introduced least squares 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 policy analysis. The distribution regression encompasses the Cox (1972) transformation/duration model as a special case, and represents a useful alternative to the Koenker and Bassett (1978) quantile regression. Second, we provide limit theory as well as inference tools for policy estimators based on quantile and distribution regression approaches. 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) and flexible duration regression (Donald, Green and Paarsch, 2000), and potential other approaches.

An alternative approach to policy 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) developed propensity score weighting estimators for counterfactual densities, while Firpo (2007) used a similar approach to construct efficient estimators of quantile treatment effects. Under correct specification, the regression and the weighting approaches are equally valid. In particular, if we use a saturated specification 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 distribution. 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 an analysis of the evolution of the U.S. wage distribution, motivated by the 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 describes our setting, the counterfactual distributions of interest, and regression models for the conditional distribution. In Section 3 we 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 additional empirical results for women and a numerical example comparing quantile and distribution regression in an online Supplementary Appendix (Chernozhukov, Fernandez-Val, and Melly, 2012).

2. The Setting and Modeling Choices for Policy Analysis

2.1. **Counterfactual distributions and policy functionals.** In order to motivate the foregoing analysis, let us first set up a simple running example. Suppose we would like to analyze the impact of gender on the marginal distribution of wages for women. Let 0 denote the population of women and 1 the population of men, Y_j denote wages, and X_j denote job market-relevant characteristics affecting wages for populations $j = 0$ and $j = 1$. The conditional distributions $F_{Y_0|X_0}$ and $F_{Y_1|X_1}$ describe the wage schedules given the observable characteristics for women and men, respectively. Let $F_{Y\langle 0|0\rangle}$ represent the observed distribution function of wages for women and $F_{Y\langle 1|0\rangle}$ represent the counterfactual distribution function of wages that would have prevailed had women faced the men's wage schedule $F_{Y_1|X_1}$. The latter distribution is called counterfactual, since it does not arise as a distribution from some 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:

$$F_{Y\langle 1|0\rangle}(y) := \int_{\mathcal{X}_0} F_{Y_1|X_1}(y|x)dF_{X_0}(x).$$

This quantity is well defined if $\mathcal{X}_1$, the support of men's characteristics, includes $\mathcal{X}_0$, the support of women's characteristics, namely $\mathcal{X}_0 \subseteq \mathcal{X}_1$. We call the difference between $F_{Y\langle 1|0\rangle}$ and $F_{Y\langle 0|0\rangle}$ the distribution *policy effect* of shifting the status quo wage schedule for women to that of men. We can also look at quantile policy effiects, the difference of quantile functions $Q_{Y\langle 1|0\rangle}$ and $Q_{Y\langle 0|0\rangle}$, as well as differences of other functionals. We stress here that the policy effects are well defined statistical parameters, and are widely used in empirical analysis. Under the conditional exogeneity assumption stated in Heckman, Lalonde, Smith (1999) and Imbens (2004), the policy effects have a causal interpretation of treatment/structural effects.

In what follows we formalize these definitions and treat 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 $Q_{Y_j|X_j}$. Thus, we can associate each F_{X_k} with label k and each $F_{Y_j|X_j}$ 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 $\mathcal{Y}_j \subseteq \mathbb{R}$.[1] 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., $\mathcal{J}\mathcal{K} = \{(j, k) : j \in \mathcal{J}, k \in \mathcal{K}\}$.

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

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:

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

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

where $F^{\leftarrow}_{Y\langle j|k\rangle}$ 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:

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

which ensures that the integral is well defined. In applications, if the support condition is not met initially, we need to explicitly trim the supports and define the parameters relative to the common support.[2]

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[3]

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

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

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

In what follows, we define a *policy* as 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 policy distribution and quantile effects

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

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

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

[2]Specifically, given initial supports $\mathcal{X}^o_j$ and $\mathcal{X}^o_k$ such that $\mathcal{X}^o_k \not\subseteq \mathcal{X}^o_j$, we can set $\mathcal{X}_k = \mathcal{X}_j = (\mathcal{X}^o_k \cap \mathcal{X}^o_j)$. Then the covariate distributions are recomputed over this support.

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

defined for non-negative outcomes only, i.e. $\mathcal{Y}_j \subseteq [0, \infty)$. In general, the policy functionals of interest 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 policy effects; Lorenz policy 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 policy effects, with $\Delta_t = G_{Y\langle j|k\rangle} - G_{Y\langle l|m\rangle}$.

2.2. **Types of policies and associated effects.** Focusing on quantile policy effects as the leading functional of interest, we can isolate the following special cases of policy effects (PE):

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

An example of type 1 PE is the gender effect on the marginal distribution of wages for women, mentioned at the beginning of the section. An example of type 2 PE is the composition effect on the change in the marginal distribution of wages over time. Concretely, let 0 denote the population of workers in 1990 and 1 denote the population of workers in 2000, Y denote wages, and X various characteristics affecting wages (age, education, experience, and other qualifications). The conditional distribution $F_{Y_j|X_j}$ describes the wage schedule given characteristics for populations $j = 0$ and $j = 1$. Then, $Q_{Y\langle 1|1\rangle}$ represents the observed quantile function of wages in 2000; $Q_{Y\langle 1|0\rangle}$ represent the counterfactual quantile function of wages in 2000, under the assumption that workers have 1990's characteristics F_{X_0} but are paid according to the 2000 wage schedule $F_{Y_1|X_1}$. The difference between the two quantile functions is the quantile PE of shifting the worker's composition in 2000 to that in 1990. Finally, we refer to Section 6 for an example of type 3 PE.

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 the second type of PE. 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 policy 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.[4] For example, this type of policy is useful for estimating the effect

[4]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

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 will cover all of them.

2.3. **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 these tasks, and our asymptotic inference theory will cover 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: $Y = P(X)'\beta + V, \quad V = Q_V(U)$, 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 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 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).[5] 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

quantiles. When the change is relatively small, the two notions coincide approximately, but generally they can differ substantially.

[5]Quantile regression is one of most important methods of regression analysis in economics. For applications, including to policy 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).

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.

Instead of restricting attention to the transformation model for the conditional distribution, we advocate to model $F_{Y|X}(y|x)$ separately for all thresholds $y \in \mathcal{Y}$, developing further the idea set forth in Foresi and Peracchi (1995).[6] Namely, we propose to consider the *distribution regression* model

$$F_{Y|X}(y|x) = \Lambda(P(x)'\beta(y)), \quad y \in \mathcal{Y}, \tag{2.8}$$

where Λ is a known link function and $\beta(\cdot)$ is an unknown functional 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, $\Lambda(v) = 1 - \exp(-\exp(v))$, $P(x)$ include a constant as the first component, and let $P(x)'\beta(y) = t(y) - P(x)'\beta$, so that the first component of $\beta(y)$ varies with the threshold y. To see the greater flexibility of (2.8), we note that (2.8) 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 $F_{Y|X}(y|x)$

[6]Foresi and Peracchi (1995) propose to estimate 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.

arbitrarily well by using rich enough $P(X)$.[7] Thus, the choice of the link function is not important for sufficiently rich $P(X)$.

Comparison. It is important to compare and contrast quantile regression and distribution regression models. Just like quantile regression generalizes location regression by allowing 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 semiparametric because they have infinite-dimensional parameters $\beta(\cdot)$. When the specification of $P(X)$ is saturated, the quantile regression and distribution regression models coincide.[8] 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.

Note that 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 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. Thus, in practice, we recommend the researchers to choose one method over the other on the basis of empirical performance, specification testing, or ability to handle complicated data situations. In section 6 we explain how these factors influence our decision in a wage regression application.

3. Estimation and Inference Methods for Policy Analysis

In this section we introduce our proposed estimation and inference methods, and outline the main estimation and inference results, without submersing into mathematical details. Note that our proposal for using distribution regressions is new for policy analysis, while our proposal for using quantile regressions builds on earlier work by Machado and Mata (2005).

3.1. Estimation of counterfactual distributions and associated policy effects.

The policy 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

[7] 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 $\lambda = \partial\Lambda$ 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 $E\left[F_{Y|X}(y|X) - \Lambda\left(P(X)'\beta\left(y\right)\right)\right]^2 \leq \bar{\lambda}\delta_p \to 0$.

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

covariate distribution $\widehat{F}_{X_k}(x)$,

$$\widehat{F}_{Y\langle j|k\rangle}(y) = \int_{\mathcal{X}_k} \widehat{F}_{Y_j|X_j}(y|x) d\widehat{F}_{X_k}(x),\ y \in \mathcal{Y}_j,\ (j,k) \in \mathcal{JK}. \tag{3.1}$$

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

$$\widehat{Q}_{Y\langle j|k\rangle}(\tau) = \widehat{F}^{r\leftarrow}_{Y\langle j|k\rangle}(\tau) \text{ and } \widehat{\Delta}_t(w) = \phi(\widehat{F}_{Y\langle j|k\rangle} : (j,k) \in \mathcal{JK})(w), \tag{3.2}$$

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).[9]

Assume that 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 $k \in \mathcal{K}_0 \subset \mathcal{K}$. 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 can have transformation samples created via transformation of some "reference" samples $k \in \mathcal{K}_0$. For example, in unconditional quantile regression, mentioned in the previous section, we create a "transformation" sample by shifting one of the covariates in the reference sample up by a unit. Formally, we let the k-th transformation sample, with $k \in \mathcal{K}_t$, be a transform of the $l(k)$-th (reference) sample, with $l(k) \in \mathcal{K}_0$: namely, $(Y_{ki}, X_{ki}) = g_{l(k),k}(Y_{l(k),i}, X_{l(k),i}), i = 1, ..., n_k$, for some transformation function $g_{l(k),k}$ and the reference indexing function $l : \mathcal{K}_t \to \mathcal{K}_0$. We also let $\mathcal{K} = \mathcal{K}_t \cup \mathcal{K}_0$.

In either case, we can estimate the covariate distribution F_{X_k} using the empirical distribution function

$$\widehat{F}_{X_k}(x) = n_k^{-1} \sum_{i=1}^{n_k} 1\{X_{ki} \le x\}, \quad k \in \mathcal{K}. \tag{3.3}$$

To estimate the conditional distribution $F_{Y_j|X_j}$, we develop methods based on the regression models described in Section 2.3. The estimator based on distribution regression (DR) takes the form:

$$\widehat{F}_{Y_j|X_j}(y|x) = \Lambda(P(x)'\widehat{\beta}_j(y)), \quad (y,x) \in \mathcal{Y}_j\mathcal{X}_j, \quad j \in \mathcal{J}, \tag{3.4}$$

$$\widehat{\beta}_j(y) = \arg\max_{b \in \mathbb{R}^p} \sum_{i=1}^{n_j} \Big[1\{Y_{ji} \le y\} \ln[\Lambda(P(X_{ji})'b)] + 1\{Y_{ji} > y\} \ln[1 - \Lambda(P(X_{ji})'b)]\Big], \tag{3.5}$$

[9] 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.

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

$$\widehat{F}_{Y_j|X_j}(y|x) = \varepsilon + \int_{\varepsilon}^{1-\varepsilon} 1\{P(x)'\widehat{\beta}_j(u) \leq y\}du, \quad (y,x) \in \mathcal{Y}_j\mathcal{X}_j, \quad j \in \mathcal{J},, \tag{3.6}$$

$$\widehat{\beta}_j(u) = \arg\min_{b\in\mathbb{R}^p} \sum_{i=1}^{n_j} [u - 1\{Y_{ji} \leq P(X_{ji})'b\}][Y_{ji} - P(X_{ji})'b], \tag{3.7}$$

for some small constant $\varepsilon > 0$. The trimming by ε is commonly employed in practice to avoid estimation of tail quantiles (Koenker, 2005, p. 148), and is valid under the conditions set forth in Theorem 4.1.[10]

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 3 allow for various additional estimators of the covariate distribution.

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

Algorithm 1 (Estimation of policy effects)**.** *(i) Obtain estimates $\widehat{F}_{X_k}$ of the covariate distributions F_{X_k} using (3.3). (ii) Apply one of the principal regression methods to obtain estimates $\widehat{F}_{Y_j|X_j}$ of the conditional distributions $F_{Y_j|X_j}$. (iii) Obtain estimates of the counterfactual distributions, quantiles and other policy functionals via (3.1) and (3.2).* □

Remark 3.1. In practice, the quantile regression coefficients can be estimated on a fine mesh $\varepsilon \leq u_1 \leq ... \leq u_S \leq 1-\varepsilon$, with meshwidth δ such that $\delta\sqrt{n_j} \to 0$. In this case the final counterfactual distribution estimator is computed as: $\widehat{F}_{Y\langle j|k\rangle}(y) = \varepsilon + n_k^{-1}\delta \sum_{i=1}^{n_k}\sum_{s=1}^{S} 1\{P(X_{ki})'\widehat{\beta}(u_s) \leq y\}$. Likewise, for distribution regression, the counterfactual distribution estimator takes the computationally convenient form $\widehat{F}_{Y\langle j|k\rangle}(y) = n_k^{-1}\sum_{i=1}^{n_k} \widehat{F}_{Y_j|X_j}(y|X_{ki})$.

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

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

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 with cross-covariance functions that depend on the type of sampling. We characterize these functions for our leading examples in Section 5, so that we can perform inference using standard analytical methods. However, for easy of inference, we recommend and prove the validity of a general resampling procedure called the *exchangeable bootstrap*. This procedure incorporates many known forms of resampling as special cases, namely the empirical

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

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. We start by defining the bootstrap weights:

Definition 1 (Exchangeable weights)**.** *For each n_k and $k \in \mathcal{K}_0$, let $(w_{k1}, ..., w_{kn_k})$ be an exchangeable, nonnegative random vector, such that for some $\epsilon > 0$*

$$\sup_{n_k} E[w_{k1}^{2+\epsilon}] < \infty, \quad {n_k}^{-1} \sum_{i=1}^{n_k} (w_{ki} - \bar{w}_k)^2 \to_{\mathbb{P}} 1, \quad \bar{w}_k^2 \to_{\mathbb{P}} 1 \geq 0, \tag{3.8}$$

where $\bar{w}_k = {n_k}^{-1} \sum_{i=1}^{n_k} w_{ki}$.[11]

Exchangeable bootstrap uses the components of the vector $(w_{k1}, ..., w_{kn_k})$ as random sampling weights in the construction of the policy estimators. In the presence of transformation samples $\mathcal{K}_t$, the exchangeable weights are inherited from the reference samples:

$$w_{ki} = w_{l(k)i}, \quad k \in \mathcal{K}_t. \tag{3.9}$$

Note that the weights constructed in this way preserve the dependence between the samples.

Remark 3.2 (Common bootstrap schemes)**.** By appropriately selecting the distribution of the weights, this procedure 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 of m_k times and 0 appears $n_k - m_k$ times ordered at random, independent of the data. □

The bootstrap version of the estimator of the counterfactual distribution is

$$\widehat{F}^*_{Y\langle j|k\rangle}(y) = \int_{\mathcal{X}_k} \widehat{F}^*_{Y_j|X_j}(y|x) d\widehat{F}^*_{X_k}(x), \quad y \in \mathcal{Y}_j, \quad (j,k) \in \mathcal{JK}. \tag{3.10}$$

[11] 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.

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 (3.3), set

$$\widehat{F}^*_{X_k}(x) = (n^*_k)^{-1} \sum_{i=1}^{n_k} w_{ki} 1\{X_{ki} \leq x\}, \quad x \in \mathcal{X}_k, \quad k \in \mathcal{K}, \tag{3.11}$$

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

$$\widehat{Q}^*_{Y\langle j|k\rangle}(\tau) = \widehat{F}^{*\leftarrow}_{Y\langle j|k\rangle}(\tau) \text{ and } \widehat{\Delta}^*_t(w) = \phi\left(\widehat{F}^*_{Y\langle j|k\rangle} : (j,k) \in \mathcal{J}\mathcal{K}\right)(w). \tag{3.12}$$

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

Algorithm 2 (Exchangeable bootstrap for a policy estimator)**.** *(i) Draw a vector of weights for the observed samples according to the definition of the exchangeable weights given above, and, if needed, construct weights for the transformation samples using (3.9). (ii) Obtain a bootstrap version $\widehat{F}^*_{X_k}$ of the covariate distribution estimator $\widehat{F}_{X_k}$ using (3.11). (iii) Obtain a 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 bootstrap versions of the estimators of the counterfactual distribution, quantiles, and other policy functionals via (3.10) and (3.12).* □

The exchangeable bootstrap distributions are useful to perform asymptotically valid inference on the policy 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

$$\widehat{F}^{\pm}_{Y\langle j|k\rangle}(y) = \widehat{F}_{Y\langle j|k\rangle}(y) \pm \widehat{t}(1-\alpha)\widehat{\Sigma}^{1/2}_{jk}(y)/\sqrt{n}, \tag{3.13}$$

such that

$$\lim_{n\to\infty} \mathbb{P}\left\{ F_{Y\langle j|k\rangle}(y) \in [\widehat{F}^{-}_{Y\langle j|k\rangle}(y), \widehat{F}^{+}_{Y\langle j|k\rangle}(y)] \text{ for all } y \in \mathcal{Y}_j \right\} = 1-\alpha. \tag{3.14}$$

Here, $\widehat{\Sigma}(y)$ is a uniformly consistent estimator of $\Sigma(y)$, the asymptotic variance 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 (3.14), we set the critical value $\widehat{t}(1-\alpha)$ as a consistent estimator of the $(1-\alpha)$-quantile of the 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 policy 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 bootstrap robust standard error estimates:* $\widehat{\Sigma}(y)^{1/2} = (q_{.75}(y) - q_{.25}(y))/1.34$ *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 (3.13) 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).

Remark 3.3. Algorithm 3 uses a robust estimator $\widehat{\Sigma}(y)$ for $\Sigma(y)$. Uniform consistency of $\widehat{\Sigma}(y)$ over $y \in \mathcal{Y}_j$ follows from the consistency of bootstrap for estimating the law of the limit Gaussian process $\bar{Z}_{jk}$ shown in Sections 4 and 5, by Lemma 1 in Chernozhukov and Fernandez-Val (2005). Uniform validity of the confidence intervals also follows from the consistency of bootstrap for estimating the law of the limit Gaussian process $\bar{Z}_{jk}$ shown in Sections 4 and 5, by the same argument as the proof of Theorem 1 in Chernozhukov and Fernandez-Val (2005), provided that $\Sigma(y)$ is bounded away from zero on the region $y \in \mathcal{Y}_j$. □

4. Inference Theory for Policy 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 $n/n_k \longrightarrow \lambda_k \in [0, \infty)$, for all $k \in \mathcal{K}$, as $n \longrightarrow \infty$.*

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. 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 function class 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 $\ell^\infty(\mathcal{Y}\mathcal{X}\mathcal{J}\mathcal{K}\mathcal{F})^2$,*

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

as stochastic processes indexed by $(y, x, j, k, f) \in \mathcal{Y}\mathcal{X}\mathcal{J}\mathcal{K}\mathcal{F}$. 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 $(j, k) \in \mathcal{J}\mathcal{K}$.

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

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

(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-neighborhood 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

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

as a stochastic process indexed by $(\tau, j, k) \in \mathcal{T}\mathcal{J}\mathcal{K}$ in the metric space $\ell^\infty(\mathcal{T}\mathcal{J}\mathcal{K})$, 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 marginal 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 policy effect estimators.

As an application of the result above, we derive functional central limit theorems for distribution and quantile policy 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 policy effects)**.** *Under the conditions of Theorem 4.1, part 1,*

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

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,

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

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 policy functionals also satisfy a functional central limit theorem. Examples include Lorenz curves and Lorenz policy effects, as well as real-valued parameters, such as Gini coefficients and Gini policy effects. Regularity conditions for Hadamard-differentiability of Lorenz and Gini functionals are given in Bhattacharya (2007).

Corollary 4.2 (Limit theory for smooth policy 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,*

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

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 bootstrap and simulation methods for policy analysis.** 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.[12] We deal with this non-pivotality by using resampling and simulation methods. An attractive result shown as part our theoretical analysis is that the policy functionals are 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 our policy estimators. This convenient result follows from the functional delta method for bootstrap of Hadamard differentiable functionals.

In order to state the results formally, 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

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

[12]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).

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 $t = (y, x, j, k, f) \in T = \mathcal{Y}\mathcal{X}\mathcal{J}\mathcal{K}\mathcal{F}$, 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 bootstrap and other simulation methods for policy estimators)**.** *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 $\sqrt{n}(\widehat{\phi}^* - \widehat{\phi}) \rightsquigarrow_{\mathbb{P}} \mathbb{Z}_\phi$ 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 policy estimators of counterfactual distribution, quantiles, distribution effects, quantile effects, and other functionals.*

This is the second main and new result of the paper. It shows that any bootstrap method is valid for estimating the limit laws of various policy functionals, 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.

5. Inference Theory for Policy 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 relabel $P(X)$ 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.**

Let us first state formally the sampling conditions introduced in Section 3.

Condition SM. *The samples $S_k = \{(Y_{ik}, X_{ik}) : 1 \leq i \leq n_k\}$, $k \in \mathcal{K}$, are generated as follows: (a) For each population $k \in \mathcal{K}_0$, S_k contains i.i.d. copies of the random vector (Y_k, X_k) that has probability law P_k. (b) For each population $k \in \mathcal{K}_t$, the samples S_k are created by transformation maps $S_k = \{g_{l(k),k}(Y_{il(k)}, X_{il(k)}) : 1 \leq i \leq n_{l(k)}\}$ for $l(k) \in \mathcal{K}_0$, as defined in Section 3. (c) Given a universal Donsker class $\mathcal{F}$, the function class $\mathcal{F} \circ g_{l(k),k}$ remains universal Donsker, which holds trivially if $g_{l(k),k}$ is an affine map or a Lipschitz map.*

Lemma D.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,

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

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 $k \in \mathcal{K}_0$,[13] and for $k \in \mathcal{K}_t$ defined by:

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

After defining the limit processes $\mathbb{G}_k$ under the two most common sampling schemes, we proceed to state the results for the leading cases formally.

5.2. **Inference theory for policy 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

$$\begin{aligned}
&\ell_{j,y,x}(Y_j, X_j) = f_{Y_j|X_j}(y|x)x'\psi_{j,F_{Y_j|X_j}(y|x)}(Y_j, X_j),\\
&\psi_{j,u}(Y_j, X_j) = -J_j(u)^{-1}\{u - 1(Y_j \leq X_j'\beta_j(u))\}X_j,\\
&\kappa_{jk,y}(Y_j, X_j, X_k) = \sqrt{\lambda_j}\int \ell_{j,y,x}(Y_j, X_j)dF_{X_k}(x) + \sqrt{\lambda_k}F_{Y_j|X_j}(y|X_k).
\end{aligned}$$

Theorem 5.1 (Validity for QR based policy 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 $\mathcal{U}_j := \{u : x'\beta(u) \in \mathcal{Y}_j$, for some $x \in \mathcal{X}_j\} \subseteq \mathcal{U}$. Then, (1) Condition D holds for the quantile regression estimator (3.6) of the conditional distribution and the empirical distribution estimator (3.3) of*

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

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 P_k-Brownian bridges defined above. 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. (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 $\bar{Z}_{jk}$ for 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 policy functionals, via Theorem 4.1 and Corollaries 4.1 and 4.2. Note that the assumption $\mathcal{U}_j \subseteq \mathcal{U}$ is the condition that permits the use of trimming in (3.6), 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 $\ell^\infty(\mathcal{U})$, where $\mathbb{G}_j$ is a P_j- Brownian Bridge. (2) The exchangeable bootstrap law is consistent for the limit law, namely, as $n_j \to \infty$,*

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

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 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). The result could be of independent interest.

5.3. **Inference Theory for Policy 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 the complementary log-log, 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). In the latter case, $y \mapsto \beta(y)$ is Lipschitz on $y \in \mathcal{Y}_j$. (c) $E\|X_j\|^2 < \infty$ and the minimum eigenvalue of

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

is bounded away from zero uniformly over $y \in \mathcal{Y}_j$.

In order to state the next result, we define

$$\begin{aligned}
&\ell_{j,y,x}(Y_j, X_j) = \lambda(x'\beta_j(y))x'\psi_{j,y}(Y_j, X_j),\\
&\psi_{j,y}(Y_j, X_j) = -J_j^{-1}(y)\frac{\Lambda(X_j'\beta_j(y)) - 1\{Y_j \le y\}}{\Lambda(X_j'\beta_j(y))(1-\Lambda(X_j'\beta_j(y)))}\lambda(X_j'\beta_j(y))X_j,\\
&\kappa_{jk,y}(Y_j, X_j, X_k) = \sqrt{\lambda_j}\int \ell_{j,y,x}(Y_j, X_j)dF_{X_k}(x) + \sqrt{\lambda_k}F_{Y_j|X_j}(y|X_k).
\end{aligned}$$

Theorem 5.2 (Validity for DR based policy 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 (3.4) of the conditional distribution and the empirical distribution estimator (3.3) of the covariate distribution, with limit processes given by*

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

where $\mathbb{G}_k$ are P_k-Brownian bridges defined above. 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. (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 $\bar{Z}_{jk}$ for 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 policy 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:

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

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

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

These limit distribution and bootstrap consistency results are new. They could be of an independent interest, and in fact they have already been applied in several studies (Chernozhukov, Fernandez-Val and Kowalski, 2011, Rothe, 2011, and Rothe and Wied, 2011). 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 illustrate our estimation and inference procedures with an analysis of 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 complement the analysis of DiNardo, Fortin, and Lemieux (1996, DFL hereafter) by providing confidence intervals for real-valued and function-valued effects of the institutional and labor market factors driving changes in the wage distribution.

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 dummy, nine education dummies interacted with experience, a quartic term in experience, two occupation dummies, twenty industry dummies, and dummies 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.[14]

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 composition 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 composition of the labor force, and (4) the price effect.

[14]Results for women can be found in Section B of the Supplementary Appendix.

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 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:

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

In constructing the decompositions (6.1), we follow the same sequential order as in DFL.[15]

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.[16] 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

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

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.

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

[15]The sequential order may matter because it defines the counterfactual distributions and the policies of interest. We report some results for the reverse sequential order in Section B of the Supplementary Appendix.

[16]We cannot identify this quantity from random variation in minimum wage, since the federal minimum wage does not vary across individuals and varies little across states in the years considered.

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

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

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

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

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 can estimate them using the plug-in principle. In particular, we estimate the conditional distribution $F_{U_0|C_0}(u|c), u \in \{0,1\}$, using logistic regression, and F_{X_1}, F_{C_1} and F_{X_0} using 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, in our application, these models are not appropriate due to substantial conditional heteroscedasticity in log wages (Lemieux, 2006, and Angrist, Chernozhukov, and Fernandez-Val, 2006). As the additional restrictions these two models impose are rejected by the data in our application, we give our preference to the distribution and quantile regression approaches.

Distribution and quantile regressions impose different parametric restrictions on the data generating process. In our application, 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, the assumptions taken from DFL imply that the wage function has different determinants below from above the minimum wage. In contrast, a 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 application 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 our application. The results presented in Section A of the Supplementary Appendix show that quantile regression is more accurate when the dependent variable is perfectly 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.[17] 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 the censoring approach. We censor wages 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 1–5. 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 1-3 refine these results by presenting estimates and 95% simultaneous confidence intervals for several major policy functionals of interest, including quantile, distribution and Lorenz policy 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 the interpretation.[18] From Figure 1, we see that the contribution of union status to the total change is quantitatively small and has a U-shaped effect across the quantile function. 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).[19] 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.

[17] Rothe and Wied (2011) suggest 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.

[18] 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.

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

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 in the upper tail of the distribution. The importance of these composition effects has been recently stressed by Lemieux (2006) and Autor, Katz and Kearney (2008). The composition effect is realized through two channels. The first channel operates through between-group inequality. In our case, more highly educated and more experienced workers earn higher wages. By increasing their proportion, we induce a larger gap between the lower and upper tails of the marginal wage distribution. The second channel is that within-group inequality varies by group, so increasing the proportion of high variance groups increases the dispersion in the marginal distribution of wages. In our case, more highly educated and more experienced workers exhibit higher within-group inequality. By increasing their proportion, we induce a higher inequality within the upper tail of the distribution. To understand the effect of these channels in 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:

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

The first channel corresponds to changes in the first term of (6.7) where $Var[X]$ represents the heterogeneity of the labor force (between group inequality); whereas the second channel corresponds to changes in the second term of (6.7) operating through the interaction of between group inequality $E[XX']$ and within group inequality $Var[\beta(U)]$.[20]

We also include estimates of the price effect. This effect captures changes in the conditional wage structure. 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 4, we check the robustness of the results with respect to the link function used to implement the distribution regression 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

[20]See Aaverge, Bjerve, and Doksum (2005) for an analogous decomposition of the pseudo-Lorenz curve.

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 distribution regression estimates of the conditional distributions below the minimum wage and recompute the functionals of interest. We find the results in Figure 5 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.

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 policy-related variables or the relationship of the outcome with these variables. The validity of the proposed inference procedures in large samples relies only on the applicability of a functional central limit theorem for the estimators of the conditional and covariate distributions. This condition holds for the most common estimators of conditional distribution and quantile functions, 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 policy 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 policy effects. Abbring and Heckman (2007) discuss this problem and various ways to identify the distribution of treatment 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.

In principle, our approach can deal with endogeneity because our high level conditions do not impose exogeneity of the regressors. In the presence of an endogenous regressor and an instrumental variable, the estimator of Chernozhukov and Hansen (2006), for instance, satisfies our technical assumptions. However, while technically covered, using an instrumental variable opens new questions about the definition of the policy and counterfactual distributions of interest, as discussed in Heckman and Vytlacil (2007a, 2007b). This problem 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

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

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 $\to_{\mathbb{P}}$ 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. A class $\mathcal{F}$ of 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).

Appendix B. Tools

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 will 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.[21] 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.

[21] 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.

Appendix C. Proof of Theorems 4.1–4.2 and Corollaries 4.1–4.2.

C.1. **Key ingredient: Hadamard differentiability of counterfactual distribution.** It will suffice 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 indices (j,k) omitted from 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, Z takes values in $UC(\mathcal{YX},\rho)$ a.s., 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 C.1 (Hadamard differentiability of counterfactual distribution)**.** *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 indicators of all 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 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 C.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 indicators of 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 $\pi^t(f) := \int f d(\Pi^t - F_X)/t$,

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

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

$$\begin{aligned}
&\int (\gamma^t(y,x) - \gamma(y,x))dF_X(x) + (\pi^t - \pi)(F_{Y|X}(y|\cdot)) + t\pi^t(\gamma(y|\cdot)) + t\pi^t(\gamma^t(y|\cdot) - \gamma(y|\cdot)) \\
&=: i(y) + ii(y) + iii(y) + iv(y).
\end{aligned}$$

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

$$\begin{aligned}
&\|iii\|_\mathcal{Y} = \left\| \int \gamma(\cdot,x) t d\pi^t(x) \right\|_\mathcal{Y} \leq \left\| \int (\gamma - \gamma \circ \pi_m)(\cdot,x) t d\pi^t(x) \right\|_\mathcal{Y} + \left\| \int (\gamma \circ \pi_m)(\cdot,x) t d\pi^t(x) \right\|_\mathcal{Y} \\
&\leq \|\gamma - \gamma \circ \pi_m\|_{\mathcal{YX}} \int |t d\pi^t| + \sum_{i=1}^m |\gamma(y_{im}, x_{im})| t\pi^t(\chi_{im}) \\
&\leq \|\gamma - \gamma \circ \pi_m\|_{\mathcal{YX}} 2 + \sum_{i=1}^m |\gamma(y_{im}, x_{im})| t\pi^t(\chi_{im}) \leq 2\epsilon + \sum_{i=1}^m |\gamma(y_{im}, x_{im})| t(\pi(\chi_{im}) + o(1)) \\
&\leq 2\epsilon + tm \left[\|\gamma\|_{\mathcal{YX}} \max_{i \leq m} \pi(\chi_{im}) + o(1) \right] \leq 2\epsilon + O(t) \to 2\epsilon,
\end{aligned}$$

since $\{\chi_{im} : 1 \leq i \leq m\} \subset \mathcal{F}$, so that $\pi^t(\chi_{im}) \to \pi(\chi_{im})$, for all $1 \leq i \leq m$. The constant ϵ is arbitrary, so the left hand side of the preceding display converges to zero.

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. □

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

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

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, it follows that

$$\begin{aligned} \sqrt{n}(\widehat{F}_{Y\langle j|k\rangle} - F_{Y\langle j|k\rangle})(y) \quad &= \sqrt{\lambda_j} \int \sqrt{n}[\widehat{F}_{Y_j|X_j}(y|x) - F_{Y_j|X_j}(y|x)] dF_{X_k}(x) \\ &\quad + \sqrt{\lambda_k} \int F_{Y_j|X_j}(y|x) \sqrt{n} d[\widehat{F}_{X_k}(x) - F_{X_k}(x)] + o_{\mathbb{P}}(1) \qquad \text{(C.1)} \\ &\rightsquigarrow \bar{Z}_{jk}(y) := \sqrt{\lambda_j} \int Z_j(y,x) dF_{X_k}(x) + \sqrt{\lambda_k} G_k(F_{Y_j|X_j}(y|\cdot)), \end{aligned}$$

jointly in $(j,k) \in \mathcal{JK}$. The first order expansion (C.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, noted in Appendix C.1. Mean square 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:

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

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

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 $[a, b]$. 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 (C.2)- (C.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. □

C.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 from by the Functional Delta Method. □

Appendix D. 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 D.1 we present a functional delta method for Z-maps (Lemma D.2) and show how to apply it to a generic Z-problem (Lemma D.3). The results of this section are of independent interest. In Section D.2 we present the proofs for Section D.1. In Section D.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 D.4 and D.5.

D.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

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

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 $\|\widehat{\Psi}(\widehat{\theta}^*(u), u)\|^2 \leq \inf_{\theta \in \Theta} \|\widehat{\Psi}^*(\theta, u)\|^2 + \widehat{r}^{*2}(u)$, 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 $r(\cdot)$-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 $\psi(\cdot, u) \in \ell^\infty(\Theta)^p$. 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 will therefore rely on Hadamard differentiability of the Z-map

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

at $(\psi, r) = (\Psi, 0)$, i.e. 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:

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 $\theta_0(u)$, 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) $\Psi(\cdot, u)$ has inverse $\Psi^{-1}(\cdot, u)$ that is continuous at 0 uniformly in $u \in \mathcal{U}$, (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 $\inf_{u \in \mathcal{U}} \inf_{\|h\|=1} \|\dot{\Psi}_{\theta_0(u), u} h\| > 0$.*

The following lemma is useful for verifying Condition Z.

Lemma D.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.*

Lemma D.2 (Hadamard differentiability of approximate Z-maps)**.** *Suppose that Condition Z 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 the derivative is defined and continuous over $z \in \ell^\infty(\Theta \times \mathcal{U})^p$.

This lemma is an alternative to Lemma 3.9.34 in VW on Hadamard differentiability of Z-functionals 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 D.2 to a generic Z-problem.

Lemma D.3 (Limit theorem for approximate Z-estimator)**.** *Suppose condition Z 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*

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

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

D.2. **Proofs of Lemma D.1-D.3. Proof of Lemma D.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). Then, we have that $\partial\theta_0(u)/\partial u = -\dot{\Psi}^{-1}_{\theta_0(u),u} \times [\partial\Psi(\theta_0(u), u)/\partial u]$ is uniformly bounded and continuous in $u \in \mathcal{U}$. 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).

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 there is $(x_t, u_t) \to (0, u)$ with $u \in \mathcal{U}$, and an element $y_t \in \Psi^{-1}(x_t, u_t) \not\to \Psi^{-1}(0, u) := \theta_0(u)$. Then, there is a further subsequence such that $y_{t_k} \to y \neq \Psi^{-1}(0, u)$ in $\overline{\mathbb{R}}^p$, and $\Psi(y_{t_k}, u_{t_k}) = x_{t_k} \to 0$. If $y_{t_k} \to y \in \mathbb{R}^p$, by continuity $\Psi(y_{t_k}, u_{t_k}) \to \Psi(y, u)$ and $\Psi(y, u) \neq 0$ since $y \neq \Psi^{-1}(0, u)$, yielding a contradiction. If $y_{t_k} \to y \in \overline{\mathbb{R}}^p \setminus \mathbb{R}^p$, we need to show that $\|\Psi(y_{t_k}, u_{t_k})\| \not\to 0$ to obtain a contradiction. Since $\|\Psi(\theta, u_{t_k})\theta/\|\theta\|\|$ is monotone in $\|\theta\|$ by $\theta \mapsto \Psi(\theta)$ being the gradient of a convex function, and is bounded above by $\|\Psi(\theta, u)\|$, it suffices to show that $\inf_{\theta \in \partial B_\delta(\theta_0(u_t))} \|\Psi(\theta, u_t)\| > c$ for some small δ. Indeed, for small enough δ, by mean-value expansion and condition (c), $\min_{\theta \in \partial B_\delta(\theta_0(u_t))} \|\Psi(\theta, u_t)\| \geq c_0 \delta > 0$.

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 $u \mapsto \theta_0(u)$. 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 $t \searrow 0$. □.

Proof of Lemma D.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 $\theta_t(u) = \phi(\psi_t, r_t)$ we need to prove that uniformly in $u \in \mathcal{U}$,

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

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 $\lambda_t(\cdot) = o(1)$.

Step 1. Here we show that uniformly in $u \in \mathcal{U}$, $\|\theta_t(u) - \theta_0(u)\| = O(t)$. Note that $\lambda_t(u) \leq \|t^{-1}\Psi(\theta_0(u), u) + z_t(\theta_0(u), u)\| = \|z(\theta_0(u), u) + o(1)\| = O(1)$ 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 that uniformly in $u \in \mathcal{U}$, $\|\Psi(\theta_t(u), u) - \Psi(\theta_0(u), u)\| = O(t)$. 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 $u \in \mathcal{U}$, $\|\theta_t(u) - \theta_0(u)\| \leq d_H(\Psi^{-1}(\Psi(\theta_t(u), u), u), \Psi^{-1}(0, u)) \to 0$, where d_H is the Hausdorff distance. By condition Z(iii) uniformly in $u \in \mathcal{U}$

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

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 will 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

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

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

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

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

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

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

Lemma D.4. *Suppose Conditions S and SM hold. Let $\mathcal{F}$ be a universal Donsker class defined on the space $\mathcal{X} \supseteq \cup_{k\in\mathcal{K}}\mathcal{X}_k$. (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 $\mathbb{G}_k(f) = \mathbb{G}_{\ell(k)}(f \circ g_{\ell(k),k})$, provided that $\mathcal{F} \circ g_{l(k),k}$ remains universal Donsker on $\mathcal{X}$.*

Proof of Lemma D.4. Statement (1) 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 the universal Donsker class given. To show Statement (2) we note that $\widehat{\mathbb{G}}_k(f) = \widehat{\mathbb{G}}_m(f \circ g_{m,k})$ for some $m = l(k)$. Recall that $l(\cdot)$ denotes the indexing function that indicates the population $l(k)$ from which the k-th population is created by transformation. Thus, $l^{-1}(m)$ is the set of all populations created from

the m-th population. Let $\mathcal{F}'$ include $\mathcal{F}$ and $\mathcal{F} \circ g_{m,k}$ for all $k \in l^{-1}(m) = \{m, ..., \} \subset \mathcal{K}$. Then $\mathcal{F}'$ is a a universal Donsker set by assumption, 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 $\mathcal{F}'$. □.

D.4. **Proof of Theorem 5.1.** (Validity of QR based Policy 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 D.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 any universal Donsker class. We use the Z-process framework described above, where we let $\theta(u) = \beta(u)$, and $\Theta = \mathbb{R}^{d_x}$. Lemma D.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 $\varphi_{u,\beta}(Y, X) = (u - 1\{Y \leq X'\beta\})X$, $\Psi(\theta, u) = P[\varphi_{u,\beta}]$, 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 $n^{1/2}\|\widehat{r}\|_{\mathcal{U}} \to_{\mathbb{P}} 0$, where ϕ is an approximate Z- map as defined in Appendix D.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 Conditions Z(i)–(iii) for $\dot{\Psi}_{\theta_0(u),u} = J(u)$. Then, by Lemma D.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 $\max_{i \leq n} w_i \|X_i\|/\sqrt{n} = o_{\mathbb{P}}(1)$, which holds since $E\|w_i X_i\|^{2+\epsilon} = E|w_i|^{2+\epsilon} E\|X_i\|^{2+\epsilon} < \infty$. 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 the entropy integral is finite and the class is F_X-Donsker for any F_X. 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 $\mathbb{R}^{d_x}$. Note that these indicators form a VC class, and hence a universal Donsker class. The final set $\mathcal{F}$ is therefore a universal Donsker class.

Next consider the mapping $\varphi : \mathbb{D}_\varphi \subset \ell^\infty(\mathcal{U})^{d_x} \mapsto \ell^\infty(\mathcal{Y}\mathcal{X})$, defined as $b \mapsto \varphi(b)$, $\varphi(b)(x, y) = \varepsilon + \int_\varepsilon^{1-\varepsilon} 1\{x'b(u) \leq y\}du$. 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

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

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

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 $(\beta, u) \mapsto (u - 1(Y \leq X'\beta))X$ 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 D.4.

Step 4. (Auxiliary: Verification of Conditions Z(i)–(iii)).We verify conditions (a)-(c) of Lemma D.1, which imply Conditions Z(i)-(iii). Conditions (a) and (b) are immediate by the assumptions. To verify (c), $\frac{\partial}{\partial(\beta', u)}\Psi(b, u) = [-E[f_{Y|X}(X'b|X)XX'], EX]$ at $(b, u) = (\beta(u), u)$, where the right side is continuous at $(b, u) = (\beta(u), u)$ for each $u \in \mathcal{U}$. This follows using the dominated convergence theorem, the a.s. continuity and boundedness of the mapping $y \mapsto f_{Y|X}(y|X)$ at $X'\beta(u)$, 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 $u \in \mathcal{U}$. □

D.5. **Proof of Theorem 5.2.** (Validity of DR based Policy 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 an 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 D.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, we know that 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 D.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 Conditions Z(i)–(iii) of Lemma D.2 for $\dot{\Psi}_{\theta_0(u),u} = J(y)$. Then, by Lemma D.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 D.3.

STEP 2.(Main: Results for conditional cdfs). Here we shall rely on compactness of $\mathcal{YX}$. 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 entropy integral is finite and the class is F_X-Donsker for any F_X. 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, and hence a universal Donsker class. The final set $\mathcal{F}$ is therefore a universal Donsker 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

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

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

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 const$\|X\|$ and envelope function 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 Z(i)–(iii)). We verify conditions (a)-(c) of Lemma D.1, which imply Conditions Z(i)-(iii). Conditions (a) and (b) are immediate by the assumptions. To verify (c), a straightforward calculation gives that at $(b, y) = (\beta(y), y)$, $\frac{\partial}{\partial(b', y)} \Psi(b, y) = [J(b, y), R(b, y)]$, where, for $H(z) = \lambda(z)/\{\Lambda(z)[1 - \Lambda(z)]\}$ and $h(z) = dH(z)/dz$,

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

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 $y \in \mathcal{Y}$. This follows from using the dominated convergence theorem, and the following ingredients: (1) a.s. continuity of the map $(b, y) \mapsto \frac{\partial}{\partial b'} \varphi_{\beta(y), y}(Y, X)$, (2) domination of $\|\frac{\partial}{\partial b'} \varphi_{b,y}(X, Y)\|$ by a square-integrable function const$\|X\|$, (3) a.s. continuity 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 $y \in \mathcal{Y}$. □

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] 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.

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

[5] Angrist, J., and J.-S. Pischke (2008): *Mostly Harmless Econometrics: An Empiricist's Companion,* Princeton Univesity Press, Princeton.
[6] Autor, D., Katz, L., and M. Kearney (2006a): "The Polarization of the U.S. Labor Market," *American Economic Review* 96, pp. 189–194.
[7] Autor, D., Katz, L., and M. Kearney (2006b): "Rising Wage Inequality: The Role of Composition and Prices," NBER Working Paper w11986.
[8] 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.
[9] 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.
[10] Bhattacharya, D. (2007): "Inference on Inequality from Household Survey Data," *Journal of Econometrics* 137(2), pp. 674–707.
[11] Buchinsky, M. (1994): "Changes in the US Wage Structure 1963-1987: Application of Quantile Regression," *Econometrica* 62, pp. 405–458.
[12] 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.
[13] 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.
[14] 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.
[15] Chernozhukov, V., and I. Fernández-Val (2005): "Subsampling Inference on Quantile Regression Processes," *Sankhyā* 67, pp. 253–276.
[16] Chernozhukov, V., I. Fernandez-Val and A. Galichon (2010): "Quantile and Probability Curves without Crossing," *Econometrica*.
[17] Chernozhukov, V., I. Fernandez-Val and A. Kowalski (2011): "Quantile Regression with Censoring and Endogeneity," CEMMAP working paper CWP20/11.
[18] Chernozhukov, V., I. Fernandez-Val and B. Melly (2009): "Inference on Counterfactual Distributions," ArXiv: 0904.0951.
[19] Chernozhukov, V., I. Fernandez-Val and B. Melly (2012): "Supplementary Appendices to Inference on Counterfactual Distributions," unpublished manuscript, MIT.
[20] Chernozhukov, V., and C. Hansen (2005): "An IV Model of Quantile Treatment Effects," *Econometrica* 73, pp. 245–261.
[21] Chernozhukov, V., and C. Hansen (2006): "Instrumental quantile regression inference for structural and treatment effect models," *Journal of Econometrics* 132, pp. 491-525.
[22] Chernozhukov, V., and H. Hong (2002): "Three-step censored quantile regression and extramarital affairs," *Journal of the American Statistical Association* 97, pp. 872–882.
[23] Cox, D. R. (1972): "Regression Models and Life Tables," (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 34, pp. 187–220.
[24] 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.
[25] 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.
[26] 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.
[27] Dudley, R. M. (1987): "Universal Donsker Classes and Metric Entropy," *The Annals of Probability* Vol. 15, No. 4, pp. 1306–1326 .
[28] Feng, X., He, X., and J. Hu (2011): "Wild bootstrap for quantile regression," *Biometrika* 98, pp. 995–999.
[29] Firpo, S. (2007): "Efficient semiparametric estimation of quantile treatment effects," *Econometrica* 75, pp. 259–276.
[30] Firpo, S., N. Fortin, and T. Lemieux (2007): "Unconditional Quantile Regressions," *Econometrica,* forthcoming.
[31] 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.
[32] Giné, E. and R. Nickl (2008): "Uniform central limit theorems for kernel density estimators," *Probability Theory and Related Fields* 141, pp. 333–387.

[33] 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.
[34] Hahn, J. (1995), "Bootstrapping Quantile Regression Estimators," *Econometric Theory* 11, pp. 105–121.
[35] Han, A., and J. A. Hausman (1990), "Flexible Parametric Estimation of Duration and Competing Risk Models," *Journal of Applied Econometrics* 5, pp. 1–28.
[36] 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.
[37] Heckman, J., R. LaLonde, and J. Smith (1999): "The economics and econometrics of active labor market programs," *Handbook of labor economics*, 3, 1865–2097.
[38] Heckman J. J. and D.. J. Vytlacil (2007a): "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.
[39] Heckman J. J. and D.. J. Vytlacil (2007b): "Econometric Evaluation of Social Programs, Part II: Using the Marginal Treatment Effect to Organize Alternative Econometric Estimators to Evaluate Social Programs, and to Forecast their Effects in New Environments," in: J.J. Heckman & E.E. Leamer (ed.), *Handbook of Econometrics*, edition 1, volume 6, chapter 71, Elsevier.
[40] 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.
[41] Horvitz, D., and D. Thompson (1952), "A Generalization of Sampling Without Replacement from a Finite Universe," *Journal of the American Statistical Association,* Vol. 47, pp. 663–685.
[42] Imbens, G. W. (2004): "Nonparametric Estimation of Average Treatment Effects Under Exogeneity: A Review," *The Review of Economics and Statistics*, 86(1), 4–29.
[43] Koenker, R. (2005), *Quantile Regression*. Econometric Society Monograph Series 38, Cambridge University Press.
[44] Koenker R. and G. Bassett (1978): "Regression Quantiles," *Econometrica* 46, pp. 33–50.
[45] Koenker, R. and J. Yoon (2009): "Parametric links for binary choice models: A Fisherian-Bayesian colloquy," *Journal of Econometrics* 152(2), pp. 120–130.
[46] Koenker, R., and Z. Xiao (2002): "Inference on the Quantile Regression Process," *Econometrica* 70, no. 4, pp. 1583–1612.
[47] Lemieux, T. (2006): "Increasing Residual Wage Inequality: Composition Effects, Noisy Data, or Rising Demand for Skill?," *American Economic Review* 96, pp. 451–498.
[48] Machado, J., and J. Mata (2005): "Counterfactual Decomposition of Changes in Wage Distributions Using Quantile Regression," *Journal of Applied Econometrics* 20, pp. 445–465.
[49] Powell, J. L. (1986): "Censored Regression Quantiles," *Journal of Econometrics* 32, pp. 143-155.
[50] Radulović, D. and M. Wegkamp (2003): "Necessary and sufficient conditions for weak convergence of smoothed empirical processes," *Statistics & Probability Letters*, 61, pp. 321–336.
[51] 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.
[52] Rothe, C. (2010): "Nonparametric Estimation of Distributional Policy Effects," *Journal of Econometrics* 155, pp. 56–70.
[53] Rothe, C. (2011): "Partial Distributional Policy Effects," unpublished manuscript , Toulouse School of Economics.
[54] Rothe, C., and D. Wied (2011): "Specification Testing in a Class of Conditional Distributional Models," unpublished manuscript, Toulouse School of Economics.
[55] Stock, J.H. (1989): "Nonparametric Policy Analysis," *Journal of the American Statistical Association* 84, pp. 567-575.
[56] 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.
[57] van der Vaart, A. (1998): *Asymptotic statistics,* Cambridge Series in Statistical and Probabilistic Mathematics, 3.
[58] van der Vaart, A., and J. Wellner (1996): *Weak convergence and empirical processes: with applications to statistics*, New York: Springer.

Table 1: Decomposing Changes in Measures of Wage Dispersion: 1979-1988

		Effect of:			
Statistic	Total change	Minimum wage	Unions	Individual attributes	Coefficients
Standard	8.0 (0.3)	2.8 (0.1)	0.7 (0.0)	1.8 (0.2)	2.7 (0.3)
Deviation		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	4.1 (0.1)	1.3 (0.0)	0.5 (0.0)	0.3 (0.1)	2.0 (0.1)
coefficient		32.1 (1.2)	11.7 (0.6)	6.8 (1.8)	49.4 (1.8)

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

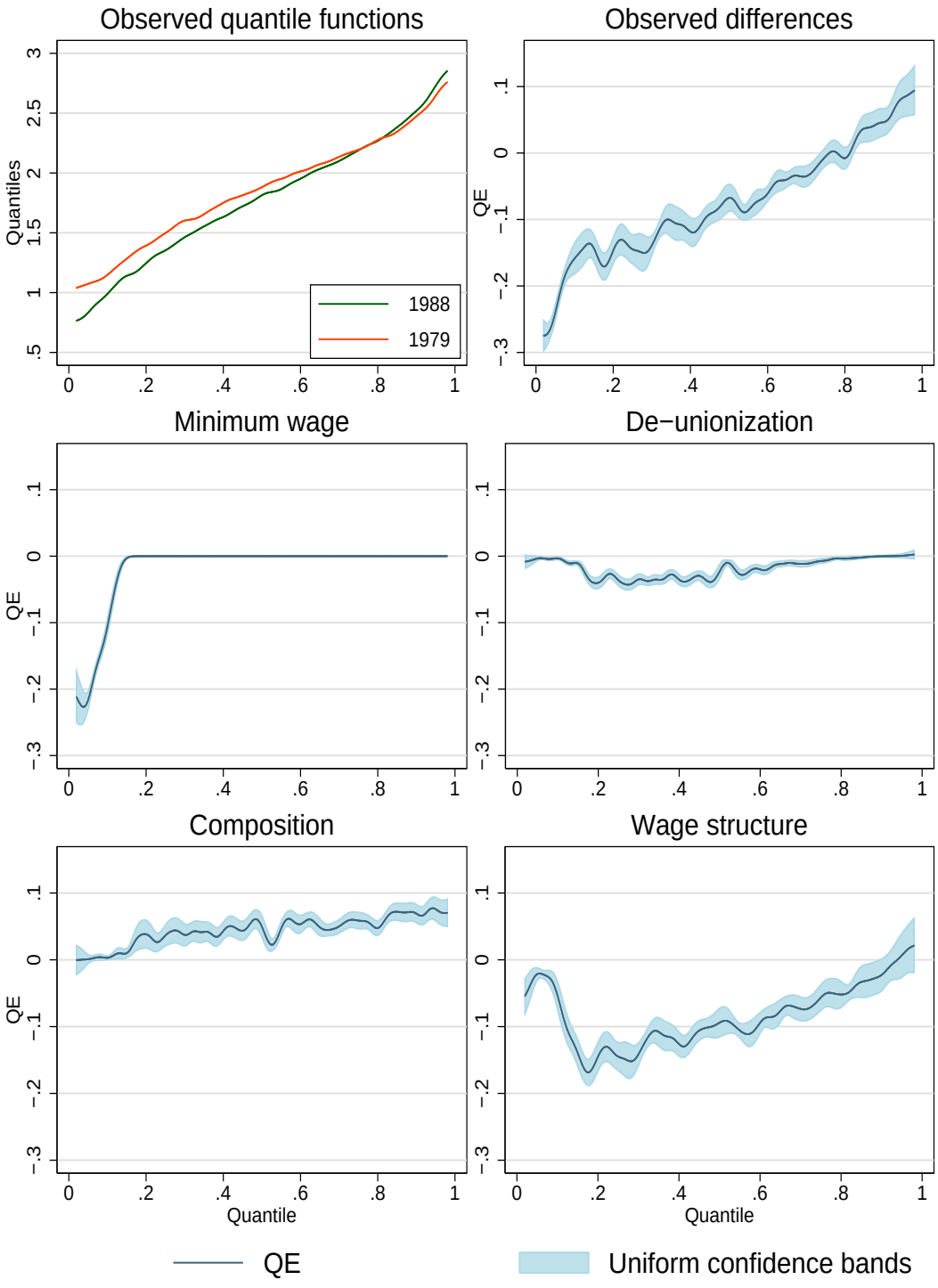

FIGURE 1. Observed quantile functions, observed differences between the quantile functions and their decomposition into four quantile policy effects. The 95% simultaneous confidence bands were obtained by empirical bootstrap with 100 repetitions. Results for men.

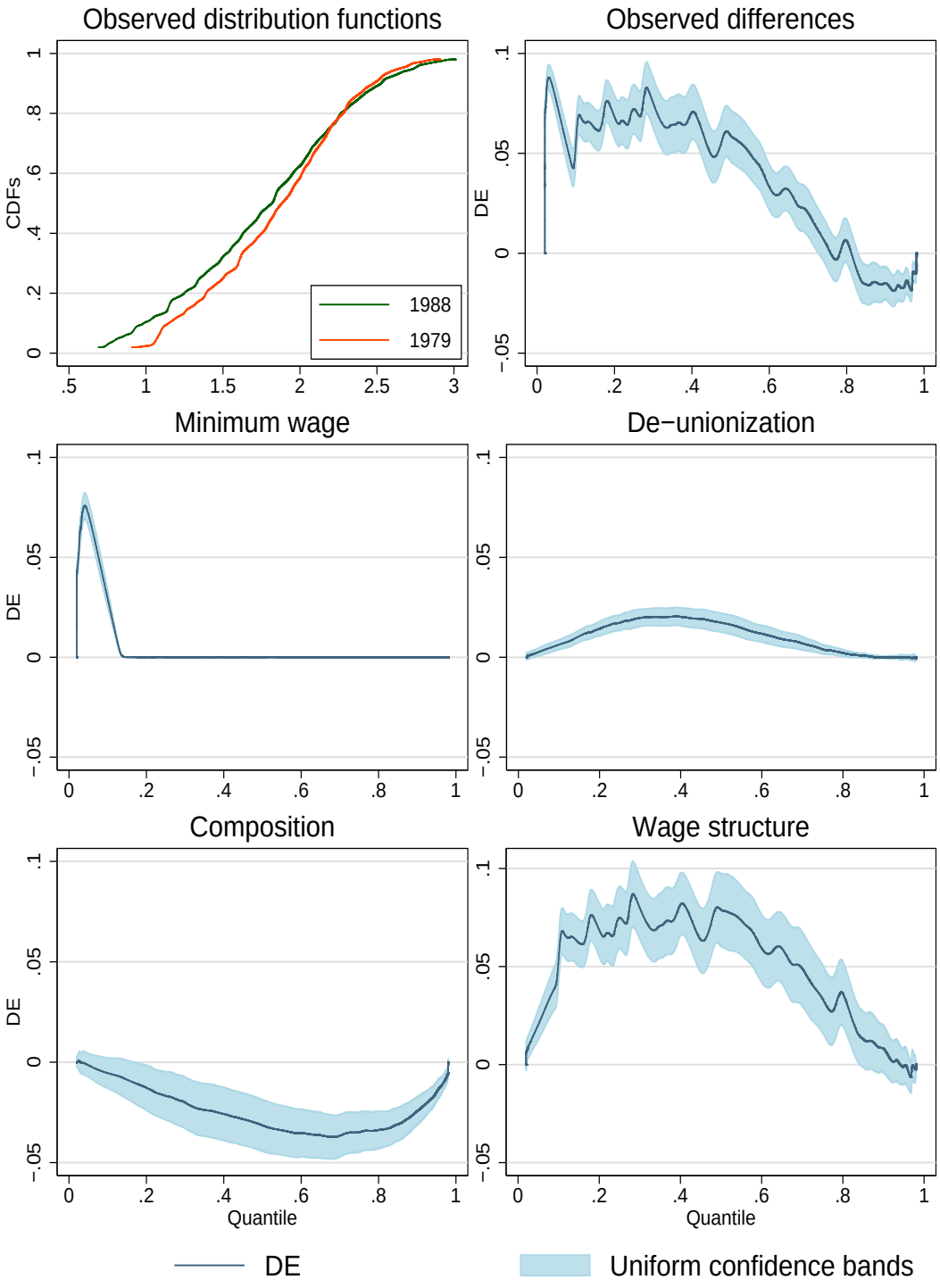

FIGURE 2. Observed distribution functions, observed differences between the distribution functions and their decomposition into four distribution policy effects. The 95% simultaneous confidence bands were obtained by empirical bootstrap with 100 repetitions. Results for men.

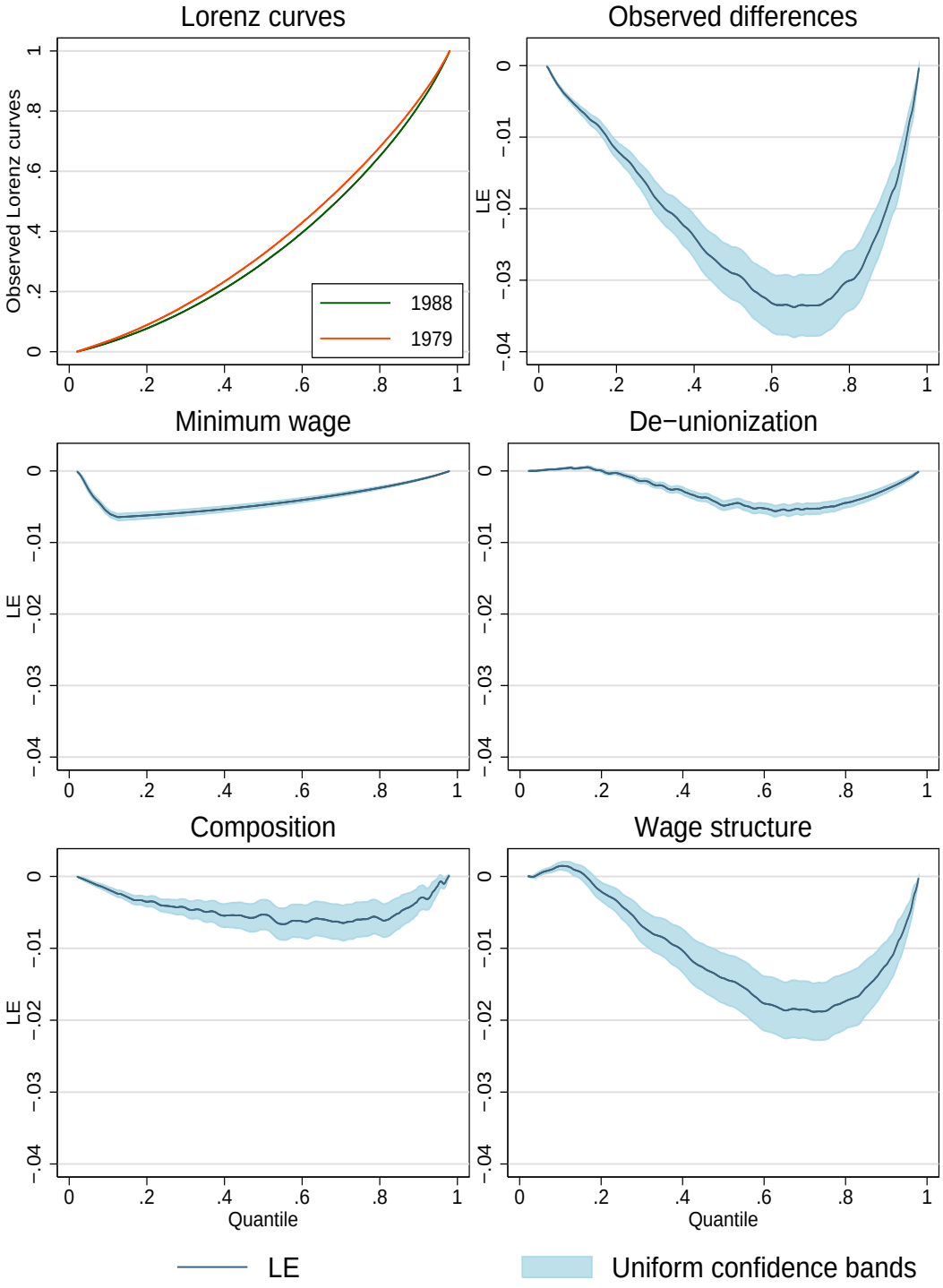

FIGURE 3. Observed Lorenz curves, observed differences between the Lorenz curves and their decomposition into four Lorenz policy effects. The 95% simultaneous confidence bands were obtained by empirical bootstrap with 100 repetitions. Results for men.

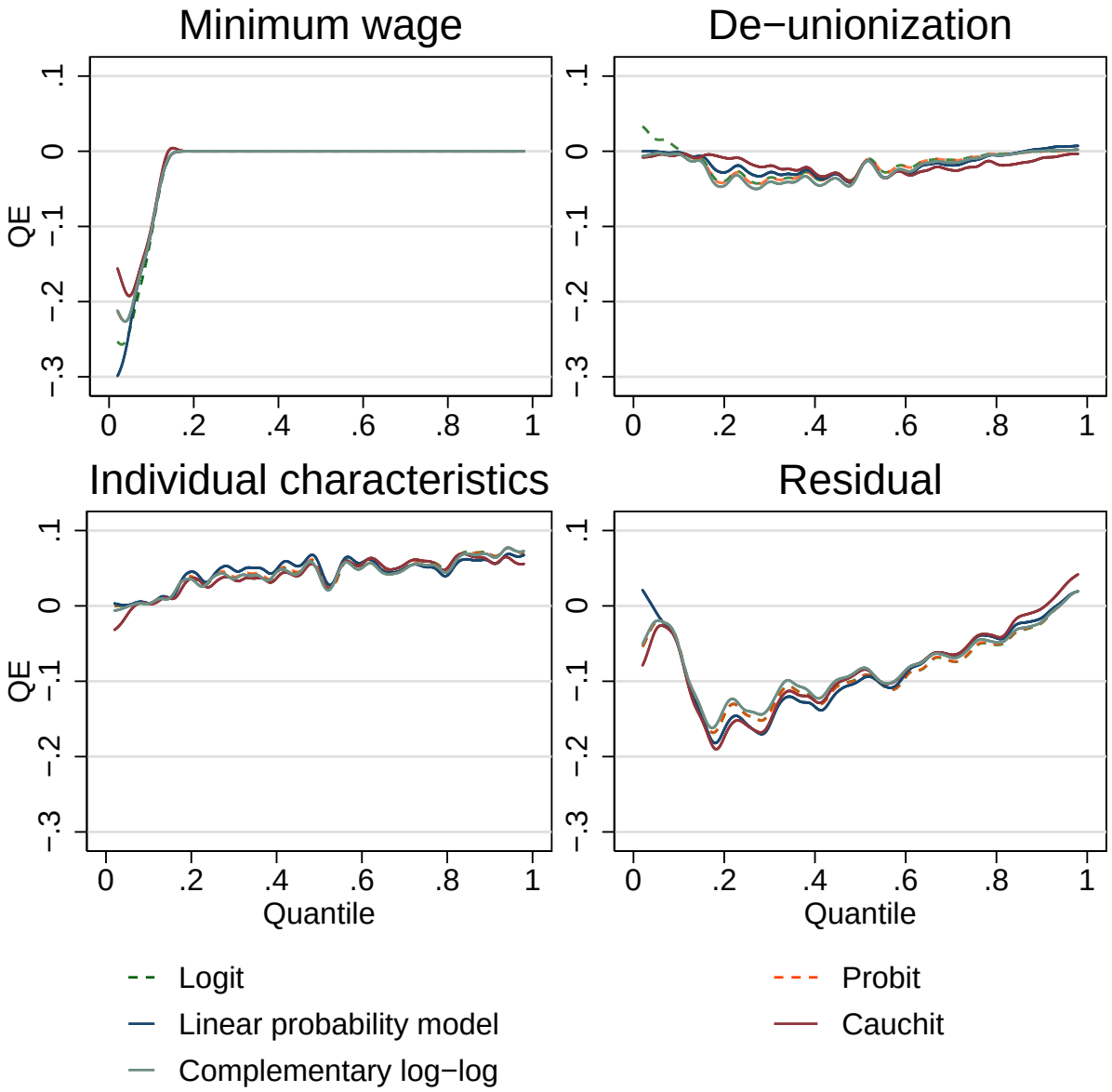

FIGURE 4. Comparison of the distribution regression estimates of the quantile policy effects based on five different link functions: logistic, normal, uniform, Cauchy and complementary log-log. Results for men.

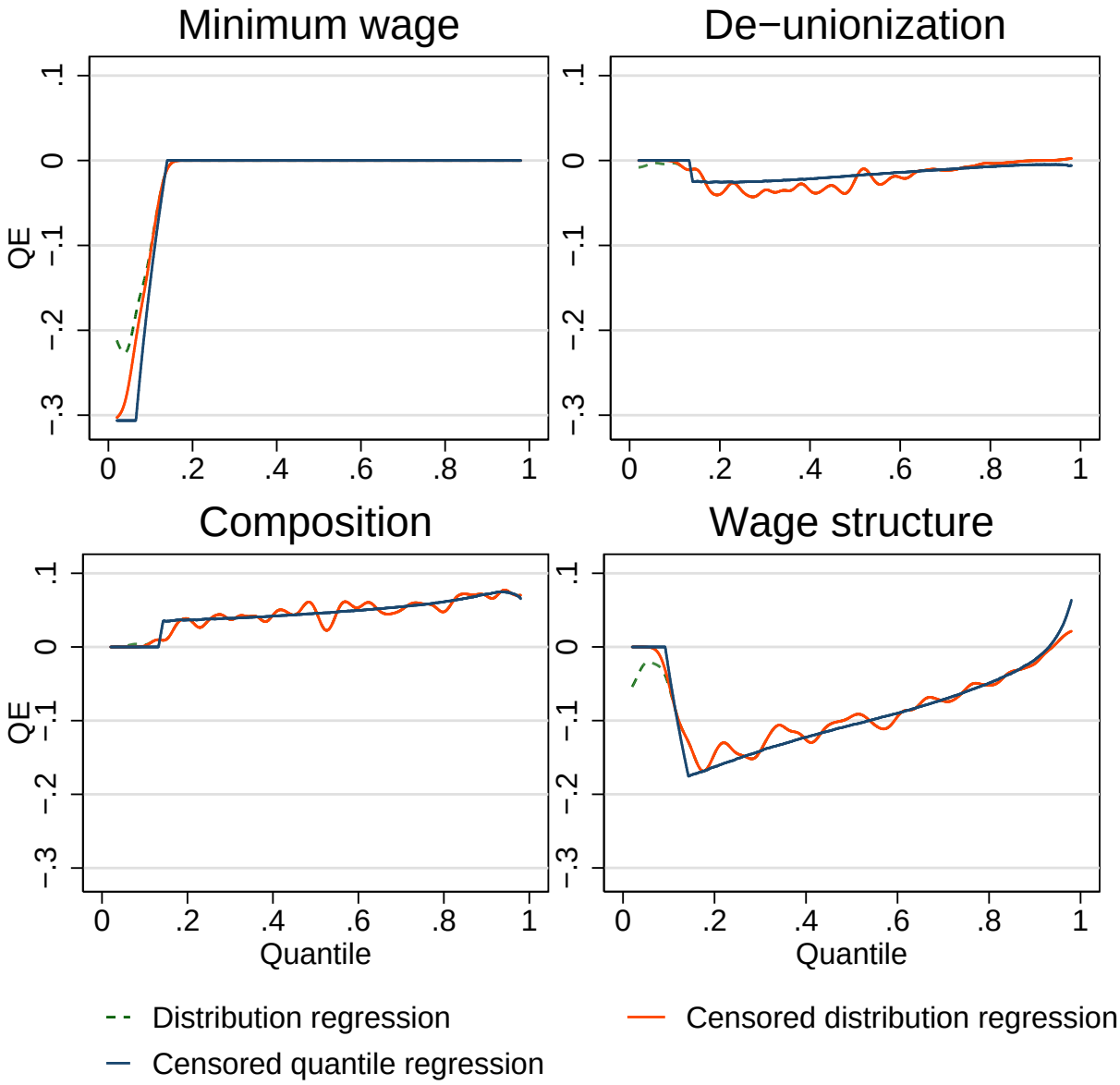

FIGURE 5. Comparison of the distribution regression, censored distribution regression and censored quantile regression estimates of the quantile policy effects. Results for men.