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

Working Paper

Inference on counterfactual distributions

cemmap working paper, No. CWP09/09

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

Suggested Citation: Chernozhukov, Victor; Fernández-Val, Iván; Melly, Blaise (2009) : Inference on counterfactual distributions, cemmap working paper, No. CWP09/09, Centre for Microdata Methods and Practice (cemmap), London, https://doi.org/10.1920/wp.cem.2009.0909

This Version is available at:
https://hdl.handle.net/10419/64778

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 CWP09/09

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

INFERENCE ON COUNTERFACTUAL DISTRIBUTIONS

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

Abstract. In this paper we develop procedures for performing inference in regression models about how potential policy interventions affect the entire marginal distribution of an outcome of interest. These policy interventions consist of either changes in the distribution of covariates related to the outcome holding the conditional distribution of the outcome given covariates fixed, or changes in the conditional distribution of the outcome given covariates holding the marginal distribution of the covariates fixed. Under either of these assumptions, we obtain uniformly consistent estimates and functional central limit theorems for the counterfactual and status quo marginal distributions of the outcome as well as other function-valued effects of the policy, including, for example, the effects of the policy on the marginal distribution function, quantile function, and other related functionals. We construct simultaneous confidence sets for these functions; these sets take into account the sampling variation in the estimation of the relationship between the outcome and covariates. Our procedures rely on, and our theory covers, all main regression approaches for modeling and estimating conditional distributions, focusing especially on classical, quantile, duration, and distribution regressions. Our procedures are general and accommodate both simple unitary changes in the values of a given covariate as well as changes in the distribution of the covariates or the conditional distribution of the outcome given covariates of general form. We apply the procedures to examine the effects of labor market institutions on the U.S. wage distribution.

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

JEL Classification: C14, C21, C41, J31, J71

Date: April 13, 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 Alberto Abadie, Josh Angrist, Manuel Arellano, David Autor, Arun Chandrasekhar, Flavio Cunha, Brigham Frandsen, Jerry Hausman, Michael Jansson, Joonhwan Lee, 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, 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.

1. Introduction

A basic objective in empirical economics is to predict the effect of a potential policy intervention or a counterfactual change in economic conditions on some outcome variable of interest. For example, we might be interested in what the wage distribution would be in 2000 if workers have the same characteristics as in 1990, what the distribution of infant birth weights would be for black mothers if they receive the same amount of prenatal care as white mothers, what the distribution of consumers expenditure would be if we change the income tax, or what the distribution of housing prices would be if we clean up a local hazardous waste site. In other examples, 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 (e.g., if female workers are paid as male workers with the same characteristics), or what the distribution of wages for black workers would be in the absence of racial discrimination in the labor market (e.g., if black workers are paid as white workers with the same characteristics). More generally, we can think of a policy intervention either as a change in the distribution of a set of explanatory variables X that determine the outcome variable of interest Y, or as a change in the conditional distribution of Y given X. Policy analysis consists of estimating the effect on the distribution of Y of a change in the distribution of X or in the conditional distribution of Y given X.

In this paper we develop procedures to perform inference in regression models about how these counterfactual policy interventions affect the entire marginal distribution of Y. The main assumption is that either the policy does not alter the conditional distribution of Y given X and only alters the marginal distribution of X, or that the policy does not alter the marginal distribution of X and only alters the conditional distribution of Y given X. Starting from estimates of the conditional distribution or quantile functions of the outcome given covariates, we obtain uniformly consistent estimates for functionals of the marginal distribution function of the outcome before and after the intervention. Examples of these functionals include distribution functions, quantile functions, quantile policy effects, distribution policy effects, means, variances, and Lorenz curves. We then construct confidence sets around these estimates that take into account the sampling variation coming from the estimation of the conditional model. These confidence sets are uniform in the sense that they cover the entire functional of interest with pre-specified probability. Our analysis specifically targets and covers the principal approaches to estimating conditional distribution models most often used in empirical work, including classical, quantile, duration, and distribution regressions. Moreover, our approach can be used to analyze the

effect of both simple interventions consisting of unitary 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 counterfactual distribution of X and conditional distribution of Y given X can correspond to known transformations of these distributions or to the distributions in a different subpopulation or group. This array of alternatives allows us to answer a wide variety of policy questions such as the ones mentioned in the first paragraph.

To develop the inference results, we establish the functional (Hadamard) differentiability of the marginal distribution functions before and after the policy with respect to the limit of the functional estimators of the conditional model of the outcome given the covariates. This result allows us to derive the asymptotic distribution for the functionals of interest taking into account the sampling variation coming from the first stage estimation of the relationship between the outcome and covariates by means of the functional delta method. Moreover, this general approach based on functional differentiability allows us to establish the validity of convenient resampling methods, such as bootstrap and other simulation methods, to make uniform inference on the functionals of interest. Because our analysis relies only on the conditional quantile estimators or conditional distribution estimators satisfying a functional central limit theorem, it applies quite broadly and we show it covers the major regression methods listed above. As a consequence, we cover a wide array of techniques, though in the discussion we devote attention primarily to the most practical and commonly used methods of estimating conditional distribution and quantile functions.

This paper contributes to the previous literature on estimating policy effects using regression methods. In particular, important developments include the work of Stock (1989), which introduced regression-based estimators to evaluate the mean effect of policy interventions, and of Gosling, Machin, and Meghir (2000) and Machado and Mata (2005), which proposed quantile regression-based policy estimators to evaluate distributional effects of policy interventions, but did not provide distribution or inference theory for these estimators. Our paper contributes to this literature by providing regression-based policy estimators to evaluate quantile, distributional, and other effects (e.g., Lorenz and Gini effects) of a general policy intervention and by deriving functional limit theory as well as practical inferential tools for these policy estimators. Our policy estimators are based on a rich variety of regression models for the conditional distribution, including classical,

quantile, duration, and distribution regressions.[1] In particular, our theory covers the previous estimators of Gosling, Machin, and Meghir (2000) and Machado and Mata (2005) as important special cases. In fact, our limit theory is generic and applies to any estimator of the conditional distribution that satisfies a functional central limit theorem. Accordingly, we cover not only a wide array of the most practical current approaches for estimating conditional distributions, but also many other existing and future approaches, including, for example, approaches that accommodate endogeneity (Abadie, Angrist, and Imbens, 2002, Chesher , 2003, Chernozhukov and Hansen, 2005, and Imbens and Newey, 2009).[2]

Our paper is also related to the literature that evaluates policy effects and treatment effects using propensity score methods. The influential article of DiNardo, Fortin, and Lemieux (1996) developed estimators for counterfactual densities using propensity score reweighting in the spirit of Horvitz and Thompson (1952). Important related work by Hirano, Imbens, and Ridder (2003) and Firpo (2007) used a similar reweighting approach in exogenous treatment effects models to construct efficient estimators of average and quantile treatment effects, respectively. As we comment later in the paper, it is possible to adapt the reweighting methods of these articles to develop policy estimators and limit theory for such estimators. Here, however, we focus on developing inferential theory for policy estimators based on regression methods, thus supporting empirical research using regression techniques as its primary method (Buchinsky, 1994, Chamberlain, 1994, Han and Hausman, 1990, Machado and Mata, 2005). The recent book of Angrist and Pischke (2008, Chap. 3) provides a nice comparative discussion of regression and propensity score methods. Finally, a related work by Firpo, Fortin, and Lemieux (2007) studied the effects of special policy interventions consisting of marginal changes in the values of the covariates. As we comment later in the paper, their approach, based on a linearization of the functionals of interest, is quite different from ours. In particular, our approach focuses on more general non-marginal changes in both the marginal distribution of covariates and conditional distribution of the outcome given covariates.

[1] We focus on semi-parametric estimators due to their dominant role in empirical work (Angrist and Pischke, 2008). In contrast, fully nonparametric estimators are practical only in situations with a small number of regressors. In future work, however, we hope to extend the analysis to nonparametric estimators.

[2] In this case, the literature provides estimators for F_{Y_d}, the distribution of potential outcome Y under treatment d, and $F_{D,Z}$, the joint distributions of (endogenously determined) treatment status D and exogenous regressors Z before and after policy. As long as the estimator of F_{Y_d} satisfies the functional central limit theorem specified in the main text and the estimator of $F_{D,Z}$ satisfies the functional central limit theorem specified in Appendix D, our inferential theory applies to the resulting policy estimators.

We illustrate our estimation and inference procedures with an analysis of the evolution of the U.S. wage distribution. Our analysis is motivated by the influential article by DiNardo, Fortin, and Lemieux (1996), which studied the institutional and labor market determinants of the changes in the wage distribution between 1979 and 1988 using data from the CPS. We complement and complete their analysis by using a wider range of techniques, including quantile regression and distribution regression, providing standard errors for the estimates of the main effects, and extending the analysis to the entire distribution using simultaneous confidence bands. Our results reinforce the importance of the decline in the real minimum wage in explaining the increase in wage inequality. They also indicate the importance of changes in both the composition of the workforce and the returns to worker characteristics in explaining the evolution of the entire wage distribution. Our results show that, after controlling for other composition effects, the process of de-unionization during the 80s played a minor role in explaining the evolution of the wage distribution.

We organize the rest of the paper as follows. In Section 2 we describe methods for performing counterfactual analysis, setting up the modeling assumptions for the counterfactual outcomes, and introduce the policy estimators. In Section 3 we derive distributional results and inferential procedures for the policy estimators. In Section 4 we present the empirical application, and in Section 5 we give a summary of the main results. In the Appendix, we include proofs and additional theoretical results.

2. Methods for Counterfactual Analysis

2.1. **Observed and counterfactual outcomes.** In our analysis it is important to distinguish between observed and counterfactual outcomes. Observed outcomes come from the population before the policy intervention, whereas (unobserved) counterfactual outcomes come from the population after the potential policy intervention. We use the observed outcomes and covariates to establish the relationship between outcome and covariates and the distribution of the covariates, which, together with either a postulated distribution of the covariates under the policy or a postulated conditional distribution of outcomes given covariates under the policy, determine the distribution of the outcome after the policy intervention, under conditions precisely stated below.

We divide our population in two groups or subpopulations indexed by $j \in \{0, 1\}$. Index 0 corresponds to the status quo or reference group, whereas index 1 corresponds to the group from which we obtain the marginal distribution of X or the conditional distribution

of Y given X to generate the counterfactual outcome distribution.[3] In order to discuss various regression models of outcomes given covariates, it is convenient to consider the following representation. Let $Q_{Y_j}(u|x)$ be the conditional u-quantile of Y given X in group j, and let F_{X_k} be the marginal distribution of the p-vector of covariates X in group k, for $j, k \in \{0, 1\}$. We can describe the observed outcome Y_j^j in group j as a function of covariates and a non-additive disturbance U_j^j via the Skorohod representation:

$$Y_j^j = Q_{Y_j}(U_j^j|X_j), \text{ where } U_j^j \sim U(0,1) \text{ independently of } X_j \sim F_{X_j}, \text{ for } j \in \{0,1\}.$$

Here the conditional quantile function plays the role of a link function. More generally we can think of $Q_{Y_j}(u|x)$ as a structural or causal function mapping the covariates and the disturbance to the outcome, where the covariate vector can include control variables to account for endogeneity. In the classical regression model, the disturbance is separable from the covariates, as in the location shift model described below, but generally it need not be. Our analysis will cover either case.

We consider two different counterfactual experiments. The first experiment consists of drawing the vector of covariates from the distribution of covariates in group 1, i.e., $X_1 \sim F_{X_1}$, while keeping the conditional quantile function as in group 0, $Q_{Y_0}(u|x)$. The counterfactual outcome Y_0^1 is therefore generated by

$$Y_0^1 := Q_{Y_0}(U_0^1|X_1), \text{ where } U_0^1 \sim U(0,1) \text{ independently of } X_1 \sim F_{X_1}. \tag{2.1}$$

This construction assumes that we can evaluate the quantile function $Q_{Y_0}(u|x)$ at each point x in the support of X_1. This requires that either the support of X_1 is a subset of the support of X_0 or we can extrapolate the quantile function outside the support of X_0.

For purposes of analysis, it is useful to distinguish two different ways of constructing the alternative distributions of the covariates. (1) The covariates before and after the policy arise from two different populations or subpopulations. These populations might correspond to different demographic groups, time periods, or geographic locations. Specific examples include the distributions of worker characteristics in different years and distributions of socioeconomic characteristics for black versus white mothers. (2) The covariates under the policy intervention arise as some known transformation of the covariates in group 0; that is $X_1 = g(X_0)$, where $g(\cdot)$ is a known function. This case covers, for

[3]Our results also cover the policy intervention of changing both the marginal distribution of X and the conditional distribution of Y given X. In this case the counterfactual outcome corresponds to the observed outcome in group 1.

example, unitary changes in the location of one of the covariates,

$$X_1 = X_0 + e_j,$$

where e_j is a unitary p-vector with a one in the position j; or mean preserving redistributions of the covariates implemented as $X_1 = (1-\alpha)E[X_0] + \alpha X_0$. These types of policies are useful for estimating the effect of smoking on the marginal distribution of infant birth weights, 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 these two cases correspond to conceptually different thought experiments, our econometric analysis will cover either situation within a unified framework.

The second experiment consists of generating the outcome from the conditional quantile function in group 1, $Q_{Y_1}(u|x)$, while keeping the marginal distributions of the covariates as in group 0, that is, $X_0 \sim F_{X_0}$. The counterfactual outcome Y_1^0 is therefore generated by

$$Y_1^0 := Q_{Y_1}(U_1^0|X_0), \text{ where } U_1^0 \sim U(0,1) \text{ independently of } X_0 \sim F_{X_0}. \tag{2.2}$$

This construction assumes that we can evaluate the quantile function $Q_{Y_1}(u|x)$ at each point x in the support of X_0. This requires that either the support of X_0 is a subset of the support of X_1 or we can extrapolate the quantile function outside the support of X_1.

In this second experiment, the conditional quantile functions before and after the policy intervention may arise from two different populations or subpopulations. These populations might correspond to different demographic groups, time periods, or geographic locations. This type of policy is useful for conceptualizing, for example, what the distribution of wages for female workers would be if they were paid as male workers with the same characteristics, or similarly for blacks or other minority groups.

We formally state the assumptions mentioned above as follows:

CONDITION M. *Counterfactual outcome variables of interest are generated by either (2.1) or (2.2). The conditional distributions of the outcome given the covariates in both groups, namely the conditional quantile functions $Q_{Y_j}(\cdot|\cdot)$ or the conditional distribution functions $F_{Y_j}(\cdot|\cdot)$ for $j \in \{0,1\}$, apply or can be extrapolated to all $x \in \mathcal{X}$, where $\mathcal{X}$ is a compact subset of $\mathbb{R}^p$ that contains the supports of X_0 and X_1.*

2.2. **Parameters of interest.** The primary (function-valued) parameters of interest are the distribution and quantile functions of the outcome before and after the policy as well as functionals derived from them.

In order to define these parameters, we first recall that the conditional distribution associated with the quantile function $Q_{Y_j}(u|x)$ is:

$$F_{Y_j}(y|x) = \int_0^1 1\left\{Q_{Y_j}(u|x) \leq y\right\} du, \quad j \in \{0,1\}. \tag{2.3}$$

Given our definitions (2.1) or (2.2) of the counterfactual outcome, the marginal distributions of interest are:

$$F^k_{Y_j}(y) := \Pr\left\{Y^k_j \leq y\right\} = \int_{\mathcal{X}} F_{Y_j}(y|x) dF_{X_k}(x), \ j,k \in \{0,1\} \tag{2.4}$$

The corresponding marginal quantile functions are:

$$Q^k_{Y_j}(u) = \inf\{y : F^k_{Y_j}(y) \geq u\}, \ j,k \in \{0,1\}.$$

The u-quantile policy effect and the y-distribution policy effect are:

$$QE^k_{Y_j}(u) = Q^k_{Y_j}(u) - Q^0_{Y_0}(u) \ \text{ and } \ DE^k_{Y_j}(y) = F^k_{Y_j}(y) - F^0_{Y_0}(y), \ j,k \in \{0,1\}.$$

It is useful to mention a couple of examples to understand the notation. For instance, $Q^1_{Y_0}(u) - Q^0_{Y_0}(u)$ is the quantile effect under a policy that changes the marginal distribution of covariates from F_{X_0} to F_{X_1}, fixing the conditional distribution of outcome to $F_{Y_0}(y|x)$. On the other hand, $Q^0_{Y_1}(u) - Q^0_{Y_0}(u)$ is the quantile effect under a policy that changes the conditional distribution of the outcome from $F_{Y_0}(y|x)$ to $F_{Y_1}(y|x)$, fixing the marginal distribution of covariates to F_{X_0}.

Other parameters of interest include, for example, Lorenz curves of the observed and counterfactual outcomes. Lorenz curves, commonly used to measure inequality, are ratios of partial means to overall means

$$L(y, F^k_{Y_j}) = \int_0^y t dF^k_{Y_j}(t) / \int_0^\infty t dF^k_{Y_j}(t),$$

defined for non-negative outcomes only. More generally, we might be interested in arbitrary functionals of the marginal distributions of the outcome before and after the interventions

$$H_Y(y) := \phi\left(y, F^0_{Y_0}, F^1_{Y_1}, F^1_{Y_0}, F^0_{Y_1}\right). \tag{2.5}$$

These functionals include the previous examples as special cases as well as other examples such as means, with $H_Y(y) = \int_{-\infty}^{\infty} t dF^k_{Y_j}(t) =: \mu^k_{Y_j}$; mean policy effects, with $H_Y(y) = \mu^k_{Y_j} - \mu^0_{Y_0}$; variances, with $H_Y(y) = \int_{-\infty}^{\infty} t^2 dF^k_{Y_j}(t) - (\mu^k_{Y_j})^2 =: (\sigma^k_{Y_j})^2$; variance policy effects, with $H_Y(y) = (\sigma^k_{Y_j})^2 - (\sigma^0_{Y_0})^2$; Lorenz policy effects, with $H_Y(y) = L(y, F^k_{Y_j}) - L(y, F^0_{Y_0}) =:$

$LE^k_{Y_j}(y)$; Gini coefficients, with $H_Y(y) = 1 - 2\int_{\mathbb{R}} L(F^k_{Y_j}, y)dy =: G^k_{Y_j}$; and Gini policy effects, with $H_Y(y) = G^k_{Y_j} - G^0_{Y_0} =: GE^k_{Y_j}$.[4]

In the case where the policy consists of either a known transformation of the covariates, $X_1 = g(X_0)$, or a change in the conditional distribution of Y given X, we can also identify the distribution and quantile functions for the effect of the policy, $\Delta^k_j = Y^k_j - Y^0_0$, by:

$$F^k_{\Delta_j}(\delta) = \int_{\mathcal{X}} \int_0^1 1\left\{Q_{\Delta_j}(u|x) \leq \delta\right\} dudF_{X_0}(x), \;\; j,k \in \{0,1\}, \tag{2.6}$$

where $Q_{\Delta_0}(u|x) = Q_{Y_0}(u|g(x)) - Q_{Y_0}(u|x)$ and $Q_{\Delta_1}(u|x) = Q_{Y_1}(u|x) - Q_{Y_0}(u|x)$; and

$$Q^k_{\Delta_j}(u) = \inf\{\delta : F^k_{\Delta_j}(\delta) \geq u\}, \;\; j,k \in \{0,1\}, \tag{2.7}$$

under the additional assumption (Heckman, Smith, and Clements, 1997):

CONDITION RP. *Conditional rank preservation:* $U^1_0 = U^0_0|X_0$ *and* $U^0_1 = U^0_0|X_0$.

2.3. **Conditional models.** The preceding analysis shows that the marginal distribution and quantile functions of interest depend on either the underlying conditional quantile function or conditional distribution function. Thus, we can proceed by modeling and estimating either of these conditional functions. We can rely on several principal approaches to carrying out these tasks. In this section we drop the dependence on the group index to simplify the notation.

Example 1. Classical regression and generalizations. Classical regression is one of the principal approaches to modeling and estimating conditional quantile functions. The classical location-shift model takes the form

$$Y = m(X) + V, \quad V = Q_V(U), \tag{2.8}$$

where $U \sim U(0,1)$ is independent of X, and $m(\cdot)$ is a location function such as the conditional mean. The disturbance V has the quantile function $Q_V(u)$, and Y therefore has conditional quantile function $Q_Y(u|x) = m(x) + Q_V(u)$. This model is parsimonious in that covariates impact the outcome only through the location. Even though this is a location model, it is clear that a general change in the distribution of covariates or the conditional quantile function can have heterogeneous effects on the entire marginal distribution of Y, affecting its various quantiles in a differential manner. The most common model for the

[4]In the rest of the discussion we keep the distribution, quantile, quantile policy effects, and distribution policy effects functions as separate cases to emphasize the importance of these functionals in practice. Lorenz curves are special cases of the general functional with $H_Y(y) = \int_{-\infty}^y tdF^k_{Y_j}(t) / \int_{-\infty}^{\infty} tdF^k_{Y_j}(t)$, and will not be considered separately.

regression function $m(x)$ is linear in parameters, $m(x) = x'\beta$, and we can estimate it using least squares or instrumental variable methods. We can leave the quantile function $Q_V(u)$ unrestricted and estimate it using the empirical quantile function of the residuals. Our results cover such common estimation schemes as special cases, since we only require the estimates to satisfy a functional central limit theorem.

The location model has played a classical role in regression analysis. Many endogenous and exogenous treatment effects models, for example, can be analyzed and estimated using variations of this model (Cameron and Trivedi, 2005 Chap. 25, and Imbens and Wooldridge, 2008). A variety of standard survival and duration models also imply (2.8) after a transformation such as the Cox model with Weibull hazard or accelerated failure time model, cf. Docksum and Gasko (1990).

The location-scale shift model is a generalization that enables the covariates to impact the conditional distribution through the scale function as well:

$$Y = m(X) + \sigma(X) \cdot V, \quad V = Q_V(U),$$

where $U \sim U(0,1)$ independently of X, and $\sigma(\cdot)$ is a positive scale function. In this model the conditional quantile function takes the form $Q_Y(u|x) = m(x) + \sigma(x)Q_V(u)$. It is clear that changes in the distribution of X or in $Q_Y(u|x)$ can have a nontrivial effect on the entire marginal distribution of Y, affecting its various quantiles in a differential manner. This model can be estimated through a variety of means (see, e.g., Rutemiller and Bowers, 1968, and Koenker and Xiao, 2002).

Example 2. Quantile regression. We can also rely on quantile regression as a principal approach to modeling and estimating conditional quantile functions. In this approach, we have the general non-separable representation

$$Y = Q_Y(U|X).$$

The model permits covariates to impact the outcome by changing not only the location and scale of the distribution but also its entire shape. An early convincing example of such effects goes back to Doksum (1974), who showed that real data can be sharply inconsistent with the location-scale shift paradigm. Quantile regression precisely addresses this issue. The leading approach to quantile regression entails approximating the conditional quantile

function by a linear form $Q_Y(u|x) = x'\beta(u)$.[5] Koenker (2005) provides an excellent review of this method.

Quantile regression allows researchers to fit parsimonious models to the entire conditional distribution. It has become an increasingly important empirical tool in applied economics. In labor economics, for example, quantile regression has been widely used to model changes in the wage distribution (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). Variations of quantile regression can be used to obtain quantile and distribution treatment effects in endogenous and exogenous treatment effects models (Abadie, Angrist, and Imbens, 2002, Chernozhukov and Hansen, 2005, and Firpo, 2007).

Example 3. Duration regression. A common way to model conditional distribution functions in duration and survival analysis is through the transformation model:

$$F_Y(y|x) = \exp(\exp(m(x) + t(y))), \tag{2.9}$$

where $t(\cdot)$ is a monotonic transformation. This model is rather rich, yet the role of covariates is limited in an important way. In particular, the model leads to the following location-shift representation:

$$t(Y) = m(X) + V,$$

where V has an extreme value distribution and is independent of X. Therefore, covariates impact a monotone transformation of the outcome only through the location function. The estimation of this model is the subject of a large and important literature (e.g., Lancaster, 1990, Donald, Green, and Paarsch, 2000, and Dabrowska, 2005).

Example 4. Distribution regression. Instead of restricting attention to transformation models for the conditional distribution, we can consider directly modeling $F_Y(y|x)$ separately for each threshold y. An example is the model

$$F_Y(y|x) = \Lambda(m(y, x)),$$

where Λ is a known link function and $m(y, x)$ is unrestricted in y. This specification includes the previous example as a special case (put $\Lambda(v) = \exp(\exp(v))$ and $m(y, x) = m(x) + t(y)$) and allows for more flexible effect of the covariates. The leading example of

[5]Throughout, by "linear" we mean specifications that are linear in the parameters but could be highly non-linear in the original covariates; that is, if the original covariate is X, then the conditional quantile function takes the form $z'\beta(u)$ where $z = f(x)$.

this specification would be a probit or logit link function Λ and $m(y,x) = x'\beta(y)$, were $\beta(y)$ is an unknown function in y (Han and Hausman, 1990, and Foresi and Peracchi, 1995). This approach is similar in spirit to quantile regression. In particular, as quantile regression, this approach leads to the specification $Y = Q_Y(U|X) = m^{-1}(\Lambda^{-1}(U), X)$ where $U \sim U(0,1)$ independently of X.

2.4. **Policy estimators and inference questions.** All of the preceding approaches generate estimates $\widehat{F}_{Y_j}(y|x),\ j \in \{0,1\}$, of the conditional distribution functions either directly or indirectly using the relation (2.3):

$$\widehat{F}_{Y_j}(y|x) = \int_0^1 1\left\{\widehat{Q}_{Y_j}(u|x) \leq y\right\} du,\ j \in \{0,1\}, \tag{2.10}$$

where $\widehat{Q}_{Y_j}(u|x)$ is a given estimate of the conditional quantile function.

We then estimate the marginal distribution functions and quantile functions for the outcome by

$$\widehat{F}^k_{Y_j}(y) = \int \widehat{F}_{Y_j}(y|x) dF_{X_k}(x), \quad \text{and} \quad \widehat{Q}^k_{Y_j}(u) = \inf\{y : \widehat{F}^k_{Y_j}(y) \geq u\},$$

respectively, for $j,k \in \{0,1\}$. We estimate the quantile and distribution policy effects by

$$\widehat{QE}^k_{Y_j}(u) = \widehat{Q}^k_{Y_j}(u) - \widehat{Q}^0_{Y_0}(u), \text{ and } \widehat{DE}^k_{Y_j}(y) = \widehat{F}^k_{Y_j}(y) - \widehat{F}^0_{Y_0}(y).$$

We estimate the general functionals introduced in (2.5) similarly, using the plug-in rule:

$$\widehat{H}_Y(y) = \phi\left(y, \widehat{F}^0_{Y_0}, \widehat{F}^1_{Y_1}, \widehat{F}^1_{Y_0}, \widehat{F}^0_{Y_1}\right). \tag{2.11}$$

For example, in this way we can construct estimates of the distribution and quantiles of the effects defined in (2.6) and (2.7).

Common inference questions that arise in policy analysis involve features of the distribution of the outcome before and after the intervention. For example, we might be interested in the average effect of the policy, or in quantile policy effects at several quantiles to measure the impact of the policy on different parts of the outcome distribution. More generally, in this analysis many questions of interest involve the entire distribution or quantile functions of the outcome. Examples include the hypotheses that the policy has no effect, that the effect is constant, or that it is positive for the entire distribution (McFadden, 1989, Barrett and Donald, 2003, Koenker and Xiao, 2002, Linton, Maasoumi, and Whang, 2005). The statistical problem is to account for the sampling variability in the estimation of the conditional model to make inference on the functionals of interests. Section 3 provides limit distribution theory for the policy estimators. This theory applies

to the entire marginal distribution and quantile functions of the outcome before and after the policy, and therefore is valid for performing either uniform inference about the entire distribution function, quantile function, or other functionals of interest, or pointwise inference about values of these functions at a specific point.

2.5. **Alternative approaches.** An alternative way to proceed with policy analysis is to use reweighting methods (DiNardo, 2002). Indeed, under Condition M, we can express the marginal distribution of the counterfactual outcome in (2.4) as

$$F^k_{Y_j}(y) = \int_{\mathcal{X}} \int_{\mathcal{Y}} 1\{Y^j_j \leq y\} w^k_j(x) dF_{Y_j}(y|x) dF_{X_j}(x), \ j,k \in \{0,1\}, \tag{2.12}$$

where $w^k_j(x) = f_{X_k}(x)/f_{X_j}(x) = (1-p_j)p_j(x)/[p_j(1-p_j(x))]$, $p_j(x) := \Pr\{J = j|X = x\}$ is the propensity score, $p_j = \Pr\{J = j\}$, J is an indicator for group j, f_{X_j} is the density of the covariate given $J = j$, and $\mathcal{Y}$ is the support of Y. The second form of the weighting function w^k_j follows from Bayes' rule. We can use the expression (2.12) along with either density or propensity score weighting to construct policy estimators. Firpo (2007) used a similar propensity score reweighting approach to derive efficient estimators of quantile effects in treatment effect models.[6] With some work, one can adapt the nice results of Firpo (2007) to obtain the results needed to perform pointwise inference, namely, inference on quantile policy effects at a specific point. However, we need to do more work to develop the results needed to perform uniform inference on the entire quantile or distribution function. We are carrying out such work in a companion paper.

In a recent important development, Firpo, Fortin, and Lemieux (2007) propose an alternative useful procedure to estimate policy effects of changes in the distribution of X. Given a functional of interest ϕ, they use a first order approximation of the policy effect:

$$\phi(F^1_{Y_0}) - \phi(F^0_{Y_0}) = \phi'(F^1_{Y_0} - F^0_{Y_0}) + R(F^1_{Y_0}, F^0_{Y_0}),$$

where $\phi'(F^1_{Y_0} - F^0_{Y_0}) = \int a(y, F^0_{Y_0}) d(F^1_{Y_0}(y) - F^0_{Y_0}(y))$ is the first order linear approximation term, where function a is the influence or the score function, and $R(F^1_{Y_0}, F^0_{Y_0})$ is the remaining approximation error. In the context of our problem, this approximation error is generally not equal to zero and does not vanish with the sample size. Firpo, Fortin, and Lemieux (2007) propose a practical mean regression method to estimate the first order term $\phi'(F^1_{Y_0} - F^0_{Y_0})$; this method cleverly exploits the law of iterated expectations and the

[6]See Angrist and Pischke (2008) for a detailed review of propensity score methods and a comparison to regression methods in the context of treatment effect models. The pros and cons of these two methods are also likely to apply to policy analysis. In this paper we focus on the regression method.

linearity of the term in the distributions. In contrast to our approach, the estimand of this method is an approximation to the policy effect with a non-vanishing approximation error, whereas we directly estimate the exact effect $\phi(F^1_{Y_0}) - \phi(F^0_{Y_0})$ without approximation error.

3. Limit Distribution and Inference Theory for Policy Estimators

In this section we provide a set of simple, general sufficient conditions that facilitate inference in large samples. We design the conditions to cover the principal practical approaches and to help us think about what is needed for various approaches to work. Even though the conditions are reasonably general, they do not exhaust all scenarios under which the main inferential methods will be valid.

3.1. Conditions on estimators of the conditional distribution and quantile functions.

We provide general assumptions about the estimators of the conditional quantile or distribution function, which allow us to derive the limit distribution for the policy estimators constructed from them. These assumptions hold for commonly used parametric and semiparametric estimators of conditional distribution and quantile functions, such as classical, quantile, duration, and distribution regressions.

We begin the analysis by stating regularity conditions for estimators of conditional quantile functions, such as classical or quantile regression. In the sequel, let $\ell^\infty((0,1)\times\mathcal{X})$ denote the space of bounded functions mapping from $(0,1)\times\mathcal{X}$ to $\mathbb{R}$, equipped with the uniform metric. We assume we have a sample $\{(X_i, Y_i), i = 1, ..., n\}$ of size n for the outcome and covariates before the policy intervention. In this sample $n_0 = n/\lambda_0$ observations come from group 0 and $n_1 = n/\lambda_1$ observations come from group 1. In what follows we use $\Rightarrow$ to denote weak convergence.

Condition C. *The conditional density $f_{Y_j}(y|x)$ of the outcome given covariates exists, and is continuous and bounded above and away from zero, uniformly on $y \in \mathcal{Y}$ and $x \in \mathcal{X}$, where $\mathcal{Y}$ is a compact subset of $\mathbb{R}$, for $j \in \{0,1\}$.*

Condition Q. *The estimators $(u,x) \mapsto \widehat{Q}_{Y_j}(u|x)$ of the conditional quantile functions $(u,x) \mapsto Q_{Y_j}(u|x)$ of outcome given covariates jointly converge in law to continuous Gaussian processes:*

$$\sqrt{n}\left(\widehat{Q}_{Y_j}(u|x) - Q_{Y_j}(u|x)\right) \Rightarrow \sqrt{\lambda_j} V_j(u,x), \ j \in \{0,1\} \tag{3.1}$$

in $\ell^\infty((0,1)\times\mathcal{X})$, *where* $(u,x)\mapsto V_j(u,x), j\in\{0,1\}$, *have zero mean and covariance function* $\Sigma_{V_{jr}}(u,x,\tilde{u},\tilde{x}) := E[V_j(u,x)V_r(\tilde{u},\tilde{x})]$, *for* $j,r\in\{0,1\}$.

These conditions appear reasonable in practice when the outcome is continuous. If the outcome is discrete, the conditions C and Q do not hold. However, in this case we can use the distribution approach discussed below. Condition C and Q focus on the case where the outcome has a compact support with a density bounded away from zero, which is a reasonable first case to analyze in detail. Condition Q applies to the most common estimators of conditional quantile functions under suitable regularity conditions (Doss and Gill, 1992, Gutenbrunner and Jureckova, 1992, Angrist, Chernozhukov, and Fernandez-Val, 2006, and Appendix F). Conditions C and Q could be extended to include other cases, without affecting subsequent results. For instance, given set $\mathcal{Y}$ in Condition C over which we want to estimate the counterfactual distribution, Condition Q needs only to hold over a smaller region $\mathcal{UX} = \{(u,x)\in(0,1)\times\mathcal{X} : Q_Y(u|x)\in\mathcal{Y}\}\subset(0,1)\times\mathcal{X}$, which leads to a less restrictive convergence requirement, without affecting any subsequent results. The joint convergence holds trivially if the samples for each group are mutually independent.

We next state regularity conditions for estimators of conditional distribution functions, such as duration or distribution regressions. Let $\ell^\infty(\mathcal{Y}\times\mathcal{X})$ denote the space of bounded functions mapping from $\mathcal{Y}\times\mathcal{X}$ to $\mathbb{R}$, equipped with the uniform metric, where $\mathcal{Y}$ is a compact subset of $\mathbb{R}$.

CONDITION D. *The estimators* $(y,x)\mapsto\widehat{F}_{Y_j}(y|x)$ *of the conditional distribution functions* $(y,x)\mapsto\widehat{F}_{Y_j}(y|x)$ *of the outcome given covariates converges in law to a continuous Gaussian processes:*

$$\sqrt{n}\left(\widehat{F}_{Y_j}(y|x)-F_{Y_j}(y|x)\right)\Rightarrow\sqrt{\lambda_j}Z_j(y,x),\ j\in\{0,1\},\tag{3.2}$$

in $\ell^\infty(\mathcal{Y}\times\mathcal{X})$, *where* $(y,x)\mapsto Z_j(y,x)$, $j\in\{0,1\}$, *have zero mean and covariance function* $\Sigma_{Z_{jr}}(y,x,\tilde{y},\tilde{x}) := E[Z_j(y,x)Z_r(\tilde{y},\tilde{x})]$, *for* $j,r\in\{0,1\}$.

This condition holds for common estimators of conditional distribution functions (Beran, 1977, Burr and Doss, 1993, and Appendix F). These estimators, however, might produce estimates that are not monotonic in the level of the outcome y (Foresi and Peracchi, 1995, and Hall, Wolff, and Yao, 1999). A way to avoid this problem and to improve the finite sample properties of the conditional distribution estimators is by rearranging the estimates (Chernozhukov, Fernandez-Val, and Galichon, 2006). The joint convergence holds trivially if the samples for each group are mutually independent.

If we start from a conditional quantile estimator $\widehat{Q}_{Y_j}(u|x)$, we can define the conditional distribution function estimator $\widehat{F}_{Y_j}(y|x)$ using the relation (2.10). It turns out that if the original quantile estimator satisfies conditions C and Q, then the resulting conditional distribution estimator satisfies condition D. This result allows us to give a unified treatment of the policy estimators based on either quantile or distribution estimators.

Lemma 1. *Under conditions C and Q, the estimators of the conditional distribution function defined by (2.10) satisfy the condition D with*

$$Z_j(y,x) = -f_{Y_j}(y|x)V_j(F_{Y_j}(y|x),x),\ j \in \{0,1\}.$$

3.2. **Examples of Conditional Estimators.** Here we verify that the principal estimators of conditional distribution and quantile functions satisfy the functional central limit theorem, which we required to hold in our main Conditions D and Q. In this section we drop the dependence on the group index to simplify the notation.

Example 1 continued. Classical regression. Consider the classical linear regression model $Y = X'\beta_0 + V$, where the disturbance V is independent of X and has mean zero, finite variance and quantile function $\alpha_0(u)$. In this case, we can estimate β_0 by mean regression and quantiles of V by the empirical quantile function of the residuals. We show in Appendix F that the resulting estimator $\widehat{\theta}(u) = (\hat{\alpha}(u), \hat{\beta}')'$ of $\theta_0(u) = (\alpha_0(u), \beta_0')'$ obeys a functional central limit theorem $\sqrt{n}(\widehat{\theta}(u) - \theta_0(u)) \Rightarrow G_0(u)^{-1}Z(u)$, where Z is a zero mean Gaussian process with covariance function $\Omega(u,\tilde{u})$ specified in (F.6) and matrix $G_0(u) := G(\alpha_0(u), \beta_0, u)'$ specified in (F.5). The resulting estimator, $\widehat{Q}_Y(u|x) = \widehat{\alpha}(u) + x'\widehat{\beta}$, of the conditional quantile function $Q_Y(u|x)$ obeys a functional central limit theorem,

$$\sqrt{n}\left(\widehat{Q}_Y(y|x) - Q_Y(y|x)\right) \Rightarrow (1,x')G_0(u)^{-1}Z(u) =: V(u,x),$$

in $\ell^\infty((0,1)\times\mathcal{X})$, where $V(u,x)$ is a zero mean Gaussian process with covariance function,

$$\Sigma_V(u,x,\tilde{u},\tilde{x}) = (1,x')G_0(u)^{-1}\Omega(u,\tilde{u})[G_0(\tilde{u})^{-1}]'(1,\tilde{x}')'.$$

Example 2 continued. Quantile regression. Consider a linear quantile regression model where $Q_Y(u|x) = x'\beta_0(u)$. In Appendix F we show the canonical quantile regression estimator satisfies a functional central limit theorem, $\sqrt{n}(\widehat{\beta}(u) - \beta_0(u)) \Rightarrow G_0(u)^{-1}Z(u)$, where $Z(u)$ is a zero mean Gaussian process with covariance function $\Omega(u,\tilde{u}) = \{\min(u,\tilde{u}) - u\cdot\tilde{u}\}E[XX']$ and $G_0(u) := G(\beta_0(u),u) = -E[f_Y(X'\beta_0(u)|X)XX']$. The estimator of the conditional quantile function also obeys a functional central limit theorem,

$$\sqrt{n}\left(\widehat{Q}_Y(u|x) - Q_Y(u|x)\right) = \sqrt{n}\left(x'\widehat{\beta}(u) - x'\beta_0(u)\right) \Rightarrow x'G_0(u)^{-1}Z(u) := V(u,x),$$

in $\ell^\infty((0,1)\times\mathcal{X})$, where $V(u,x)$ is a zero mean Gaussian process with covariance function given by:

$$\Sigma_V(u,x,\tilde{u},\tilde{x}) = x'G_0(u)^{-1}\Omega(u,\tilde{u})G_0(\tilde{u})^{-1}\tilde{x}.$$

Example 3 continued. Duration regression. Consider the transformation model for the conditional distribution function stated in equation (2.9). A common duration model that gives rise to this specification is the proportional hazard model of Cox (1972), where the conditional hazard rate of an individual with covariate vector x is $\lambda_Y(y|x) = \lambda_0(y)\exp(x'\beta_0)$, β_0 is a p-vector of regression coefficients, λ_0 is the nonnegative baseline hazard rate function, and $y \in \mathcal{Y} = [0,\bar{y}]$ for some maximum duration $\bar{y}$. Let $\Lambda_0(y) = \int_0^y \lambda_0(\tilde{y})d\tilde{y}$ denote the integrated baseline hazard function. Then $F_Y(y|x) = 1 - \exp\{-\exp(x'\beta_0 + \ln\Lambda_0(y))\}$, delivering the transformation model (2.9) with $t(y) = \ln\Lambda_0(y)$ and $m(x) = x'\beta_0$.

In order to discuss estimation, let us assume i.i.d. sampling of (Y_i, X_i) without censoring. Then Cox's (1972) partial maximum likelihood estimator of β_0 takes the form

$$\widehat{\beta} = \arg\max_\beta \int \sum_{i=1}^n \log\Big\{J_i(y)\exp(x_i'\beta)/\sum_{j=1}^n J_j(y)\exp(x_j'\beta)\Big\}dN_i(y),$$

and the Breslow-Nelson-Aalen estimator of Λ_0 takes the form

$$\widehat{\Lambda}(y) = \int_0^y \Big\{\sum_{j=1}^n J_j(\tilde{y})\exp(x_j'\widehat{\beta})\Big\}^{-1} d\Big\{\sum_{i=1}^n N_i(\tilde{y})\Big\},$$

where $N_i(y) := 1\{Y_i \le y\}$ and $J_i(y) := 1\{Y_i \ge y\}$, $y \in \mathcal{Y}$; see Breslow (1972,1974).

Let W denote a standard Brownian motion on $\mathcal{Y}$ and let Z denote an independent p-dimensional standard normal vector. Andersen and Gill (1982) show that

$$\sqrt{n}(\widehat{\beta}-\beta_0, \widehat{\Lambda}(y)-\Lambda_0(y)) \Rightarrow (\Sigma^{-1/2}Z, W(a(y)) - b(y)'\Sigma^{-1/2}Z)$$

in $\mathbb{R}^p \times \ell^\infty(\mathcal{Y})$, with the terms $a(y)$, $b(y)$, and Σ, and regularity conditions defined in Andersen and Gill (1982) and Burr and Doss (1993). Let $\widehat{F}_Y(y|x) = 1 - \exp\{-\exp(x'\widehat{\beta} + \log\widehat{\Lambda}(y))\}$ be the estimator of $F_Y(y|x)$. Since $F_Y(y|x)$ is Hadamard-differentiable in (β,Λ), by the functional delta method we have the functional central limit theorem

$$\sqrt{n}\left(\widehat{F}_Y(y|x) - F_Y(y|x)\right) \Rightarrow \{1-F_Y(y|x)\}\left\{\exp(x'\beta_0)W(a(y)) + b(y,x)'\Sigma^{-1/2}Z\right\} =: Z(y,x),$$

in $\ell^\infty(\mathcal{Y}\times\mathcal{X})$, where $b(y,x) = \lambda_Y(y|x)x - \exp(x'\beta_0)b(y)$, and $Z(y,x)$ is a zero mean Gaussian process with covariance function, for $y \le \tilde{y}$,

$$\Sigma_Z(y,x,\tilde{y},\tilde{x}) = \{1-F_Y(y|x)\}\{1-F_Y(\tilde{y}|\tilde{x})\}\left\{\exp(x'\beta_0)\exp(\tilde{x}'\beta_0)a(y) + b(y,x)'\Sigma^{-1}b(\tilde{y},\tilde{x})\right\}.$$

In Appendix F we also discuss another estimator of this model.

Example 4 continued. Distribution regression. Consider the model $F_Y(y|x) = \Lambda(x'\beta_0(y))$ for the conditional distribution function, where Λ is a known link function, such as the logistic or normal distribution. We can estimate the function $\beta_0(y)$ by applying maximum likelihood to the indicator variables $1\{Y \leq y\}$ for each value of $y \in \mathcal{Y}$ separately. In Appendix F, we prove that the resulting estimator $\widehat{\beta}(y)$ of $\beta_0(y)$ obeys a functional central limit theorem

$$\sqrt{n}\left(\widehat{\beta}(y) - \beta_0(y)\right) \Rightarrow -G_0(y)^{-1}Z(y),$$

where $G_0(y) := G(\beta_0(y), y) = E[\lambda[X'\beta_0(y)]^2 XX'/\{\Lambda[X'\beta_0(y)](1 - \Lambda[X'\beta_0(y)])\}]$, λ is the derivative of Λ, and $Z(y)$ is a zero mean Gaussian process with covariance function

$$\Omega(y, \tilde{y}) = E\left[XX'\lambda[X'\beta_0(y)]\lambda[X'\beta_0(\tilde{y})]/\{\Lambda[X'\beta_0(y)](1 - \Lambda[X'\beta_0(\tilde{y})])\}\right],$$

for $\tilde{y} \geq y$. Hence the resulting estimator $\widehat{F}_Y(y|x) := \Lambda(x'\widehat{\beta}(y))$ of the conditional distribution function also obeys the functional central limit theorem,

$$\sqrt{n}\left(\widehat{F}_Y(y|x) - F_Y(y|x)\right) \Rightarrow -\lambda[x'\beta_0(y)]x'G_0(y)^{-1}Z(y) =: Z(y, x),$$

in $\ell^\infty(\mathcal{Y} \times \mathcal{X})$, where $Z(y, x)$ is a zero mean Gaussian process with covariance function:

$$\Sigma_Z(y, x, \tilde{y}, \tilde{x}) = \lambda[x'\beta_0(y)]\lambda[\tilde{x}'\beta_0(\tilde{y})]x'G_0(y)^{-1}\Omega(y, \tilde{y})G_0(\tilde{y})^{-1}\tilde{x}.$$

3.3. **Basic principles underlying the limit theory.** The derivation of the limit theory for policy estimators relies on several basic principles that allow us to link the properties of the estimators of conditional (quantile and distribution) functions with the properties of estimators of marginal functions. First, although there does not exist a direct connection between conditional and marginal quantiles, we can always switch from conditional quantiles to conditional distributions using Lemma 1, then use the law of iterated expectations to go from conditional distribution to marginal distribution, and finally get to marginal quantiles by inverting. Second, as the functionals of interest depend on the entire conditional function, we must rely on the functional delta method to obtain the limit theory for these functionals as well as to obtain intermediate limit results such as Lemma 1. Since the estimated conditional distributions and quantile functions are usually non-monotone and discontinuous in finite samples, we must use refined forms of the functional delta method.

Accordingly, the key ingredient in the derivation and one of the main theoretical contributions of the paper is the demonstration of the Hadamard differentiability of the functionals of interest with respect to the limit of the conditional processes, tangentially to the subspace of continuous functions. Indeed, we need this refined form of differentiability to

deal with our conditional processes, which typically are discontinuous random functions in finite samples yet converge to continuous random functions in large samples. These refined differentiability results in turn enable us to use the functional delta method to derive all of the following limit distribution and inference theory.

3.4. **Limit theory for counterfactual distribution and quantile functions.** Our first main result shows that the estimators of the marginal distribution and quantile functions before and after the policy intervention satisfy a functional central limit theorem.

Theorem 1 (Limit distribution for marginal distribution functions)**.** *Under Conditions M and D, the estimators $\widehat{F}^k_{Y_j}(y)$ of the marginal distribution functions $F^k_{Y_j}(y)$ jointly converge in law to the following Gaussian processes:*

$$\sqrt{n}\left(\widehat{F}^k_{Y_j}(y) - F^k_{Y_j}(y)\right) \Rightarrow \sqrt{\lambda_j}\int_{\mathcal{X}} Z_j(y,x)dF_{X_k}(x) =: \sqrt{\lambda_j}Z^k_j(y), \quad j,k \in \{0,1\}, \tag{3.3}$$

in $\ell^\infty(\mathcal{Y})$, where $y \mapsto Z^k_j(y)$, $j \in \{0,1\}$, have zero mean and covariance function, for $j,k,r,s \in \{0,1\}$,

$$\Sigma^{ks}_{Z_{jr}}(y,\tilde{y}) := E[Z^k_j(y)Z^s_r(\tilde{y})] = \int_{\mathcal{X}}\int_{\mathcal{X}} \Sigma_{Z_{jr}}(y,x,\tilde{y},\tilde{x})dF_{X_k}(x)dF_{X_s}(\tilde{x}). \tag{3.4}$$

Theorem 2 (Limit distribution for marginal quantile functions)**.** *Under Conditions M, C, and D the estimators $\widehat{Q}^k_{Y_j}(u)$ of the marginal quantile functions $Q^k_{Y_j}(u)$ jointly converge in law to the following Gaussian processes:*

$$\sqrt{n}\left(\widehat{Q}^k_{Y_j}(u) - Q^k_{Y_j}(u)\right) \Rightarrow -Z^k_j(Q^k_{Y_j}(u))/f^k_{Y_j}(Q^k_{Y_j}(u)) =: V^k_j(u), \quad j,k \in \{0,1\}, \tag{3.5}$$

in $\ell^\infty((0,1))$, where $f^k_{Y_j}(y) = \int_{\mathcal{X}} f_{Y_j}(y|x)dF_{X_k}(x)$, and $u \mapsto V^k_j(u)$, $j,k \in \{0,1\}$, have zero mean and covariance function, for $j,k,r,s \in \{0,1\}$,

$$\Sigma^{ks}_{V_{jr}}(u,\tilde{u}) := E[V^k_j(u)V^s_r(\tilde{u})] = \Sigma^{ks}_{Z_{jr}}(Q^k_{Y_j}(u), Q^s_{Y_r}(\tilde{u}))/[f^k_{Y_j}(Q^k_{Y_j}(u))f^s_{Y_r}(Q^s_{Y_r}(\tilde{u}))].$$

Our second main result shows that the estimators of the marginal quantile and distribution policy effects also satisfy a functional central limit theorem.

Corollary 1 (Limit distribution for quantile policy effects)**.** *Under Conditions M, C, and D the estimators of the quantile policy effects converge in law to the following Gaussian processes:*

$$\sqrt{n}\left(\widehat{QE}^k_{Y_j}(u) - QE^k_{Y_j}(u)\right) \Rightarrow \sqrt{\lambda_j}V^k_j(u) - \sqrt{\lambda_0}V^0_0(u) =: W^k_j(u), \ k,j \in \{0,1\}, \tag{3.6}$$

in the space $\ell^\infty((0,1))$, where the processes $u \mapsto W^k_j(u)$, $j,k \in \{0,1\}$, have zero mean and covariance function $\Sigma^{ks}_{W_{jr}}(u,\tilde{u}) := E[W^k_j(u)W^s_r(\tilde{u})]$, for $j,k,r,s \in \{0,1\}$.

Corollary 2 (Limit distribution for distribution policy effects)**.** *Under Conditions M and D the estimators of the distribution policy effects converge in law to the following Gaussian processes:*

$$\sqrt{n}\left(\widehat{DE}^{k}_{Y_j}(y) - DE^{k}_{Y_j}(y)\right) \Rightarrow \sqrt{\lambda_j}Z^{k}_{j}(y) - \sqrt{\lambda_0}Z^{0}_{0}(y) =: S^{k}_{j}(y),\ j,k \in \{0,1\}, \tag{3.7}$$

in the space $\ell^\infty(\mathcal{Y})$, where the processes $y \mapsto S^k_j(y)$, $j,k \in \{0,1\}$, have zero mean and variance function $\Sigma^{ks}_{S_{jr}}(y,\tilde{y}) := E[S^k_j(y)S^s_r(\tilde{y})]$, for $j,k,r,s \in \{0,1\}$.

Our third main result shows that various functionals of the status quo and counterfactual marginal distribution and quantile functions satisfy a functional central limit theorem.

Corollary 3 (Limit distribution for differentiable functionals)**.** *Let $H_Y(y) = \phi(y, F^0_{Y_0}, F^1_{Y_1}, F^1_{Y_0}, F^0_{Y_1})$, a functional taking values in $\ell^\infty(\mathcal{Y})$, be Hadamard differentiable in $(F^0_{Y_0}, F^1_{Y_1}, F^1_{Y_0}, F^0_{Y_1})$ tangentially to the subspace of continuous functions with derivative $(\phi'_{00}, \phi'_{11}, \phi'_{01}, \phi'_{10})$. Then under Conditions M and D the plug-in estimator $\widehat{H}_Y(y)$ defined in (2.11) converges in law to the following Gaussian process:*

$$\sqrt{n}\left(\widehat{H}_Y(y) - H_Y(y)\right) \Rightarrow \sum_{j,k\in\{0,1\}} \sqrt{\lambda_j}\phi'_{jk}(y, F^0_{Y_0}, F^1_{Y_1}, F^1_{Y_0}, F^0_{Y_1})Z^k_j(y) =: T_H(y), \tag{3.8}$$

in $\ell^\infty(\mathcal{Y})$, where $y \mapsto T_H(y)$ has zero mean and covariance function $\Sigma_{T_H}(y,\tilde{y}) := E[T_H(y)T_H(\tilde{y})]$.

Examples of functionals covered by Corollary 3 include function-valued parameters, such as Lorenz curves and Lorenz policy effects, as well as scalar-valued parameters, such as Gini coefficients and Gini policy effects (Barrett and Donald, 2009). These examples also include quantile and distribution functions of the effect of the policy defined under Condition RP; in Appendix C we state the results for these effects separately in order to give them some emphasis.

3.5. **Uniform inference and resampling methods.** We can readily apply the preceding limit distribution results to perform inference on the distributions and quantiles of the outcome before and after the policy at a specific point. For example, Corollary 1 implies that the quantile policy effect estimator for a given quantile u is asymptotically normal with mean $QE^k_{Y_j}(u)$ and variance $\Sigma^{kk}_{W_{jj}}(u,u)/n$. We can therefore perform inference on $QE^k_{Y_j}(u)$ for a particular quantile index u using this normal distribution and replacing $\Sigma^{kk}_{W_{jj}}(u,u)$ by a consistent estimate.

However, pointwise inference permits looking at the effect of the policy at a specific point only. This approach might be restrictive for policy analysis where the quantities and hypotheses of interest usually involve many points or a continuum of points. That is, the entire distribution or quantile function of the observed and counterfactual outcomes is often of interest. For example, in order to test hypotheses of the policy having no effect on the distribution, having a constant effect throughout the distribution, or having a first order dominance effect, we must use the entire outcome distribution, and not only a single specific point. Moreover, simultaneous inference corrections to pointwise procedures based on the normal distribution, such as Bonferroni-type corrections, can be very conservative for simultaneous testing of highly dependent hypotheses, and become completely inadequate for testing a continuum of hypotheses.

A convenient and computationally attractive approach for performing inference on function-valued parameters is to use Kolmogorov-Smirnov type procedures. Some complications arise in our case because the limit processes are non-pivotal, as their covariance functions depend on unknown, though estimable, nuisance parameters.[7] A practical and valid way to deal with non-pivotality is to use resampling and related simulation methods. An attractive feature of our theoretical analysis is that validity of resampling and simulation methods follows from the Hadamard differentiability of the policy functionals with respect to the underlying conditional functions. Indeed, given that bootstrap and other methods can consistently estimate the limit laws of the estimators of the conditional distribution and quantile functions, they also consistently estimate the limit laws of our policy estimators. This convenient result follows from preservation of validity of bootstrap and other resampling methods for estimating laws of Hadamard differentiable functionals; see more on this in Lemma 6 in Appendix A.

Theorem 3 (Validity of bootstrap and other simulation methods for estimating the laws of policy estimators of function-valued parameters)**.** *If the bootstrap or any other simulation method consistently estimates the laws of the limit stochastic processes (3.1) and (3.2) for the estimators of the conditional quantile or distribution function, then this method also consistently estimates the laws of the limit stochastic processes (3.3), (3.5), (3.6), (3.7), and (3.8) for policy estimators of marginal distribution and quantile functions and other functionals.*

[7]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).

Theorem 3 shows that the bootstrap is valid for estimating the limit laws of various inferential processes. This is true provided that the bootstrap is valid for estimating the limit laws of the (function-valued) estimators of the conditional distribution and quantile functions. This is a reasonable condition, but, to the best of our knowledge, there are no results in the literature that verify this condition for our principal estimators. Indeed, the previous results on the bootstrap established its validity only for estimating the pointwise laws of our principal estimators, which is not sufficient for our purposes.[8] To overcome this difficulty, in Appendix F we prove validity of the empirical bootstrap and other related methods, such as Bayesian bootstrap, wild bootstrap, k out of n bootstrap, and subsampling bootstrap, for estimating the laws of function-valued estimators, such as quantile regression and distribution regression processes. These results may be of substantial independent interest.

We can then use Theorem 3 to construct the usual uniform bands and perform inference on the marginal distribution and quantile functions, and various functionals, as described in detail in Chernozhukov and Fernandez-Val (2005) and Angrist, Chernozhukov, and Fernandez-Val (2006). Moreover, if the sample size is large, we can reduce the computational complexity of the inference procedure by resampling the first order approximation to the estimators of the conditional distribution and quantile functions (Chernozhukov and Hansen, 2006); by using subsampling bootstrap (Chernozhukov and Fernandez-Val, 2005); or by simulating the limit processes Z_j or V_j, $j \in \{0, 1\}$, appearing in expressions (3.1) and (3.2), using multiplier methods (Barrett and Donald, 2003).

3.6. **Incorporating uncertainty about the distribution of the covariates.** In the preceding analysis we assumed that we know the distributions of the covariates before and after the policy intervention for the target population. In practice, however, we usually observe such distributions only for individuals in the sample. If the individuals in the sample are the target population, then the previous limit theory is valid for performing inference without any adjustments. If a more general population group is the target population, then the distributions of the covariates need to be estimated, and the previous limit theory needs to be adjusted to take this into account. Here we highlight the main ideas, while in Appendix D we present formal distribution and inference theory.

We begin by assuming that the estimators $x \mapsto \widehat{F}_{X_k}(x)$, $k \in \{0, 1\}$, of the covariate distribution functions are well behaved, specifically that they converge jointly in law to

[8]Exceptions include Chernozhukov and Hansen (2006) and Chernozhukov and Fernandez-Val (2005), but they looked at forms of subsampling only.

Gaussian processes $B_{X_k}, k \in \{0,1\}$:

$$\sqrt{n}\left(\widehat{F}_{X_k}(x) - F_{X_k}(x)\right) \Rightarrow \sqrt{\lambda_j} B_{X_k}(x), \quad k \in \{0,1\},$$

as rigorously defined in Appendix D.1. This assumption is quite general and holds for conventional estimators such as the empirical distribution under i.i.d. sampling as well as various modifications of conventional estimators, as discussed further in Appendix D. The joint convergence holds trivially in the leading cases where the distribution in group 1 is a known transformation of the distribution in group 0, or when the two distributions are estimated from independent samples.

The estimation of the covariate distributions affects limit distributions of functionals of interests. Let us consider, for example, the marginal distribution functions. When the covariate distributions are unknown, the plug-in estimators for these functions take the form $\widehat{F}^k_{Y_j}(y) = \int_{\mathcal{X}} \widehat{F}_{Y_j}(y|x) d\widehat{F}_{X_k}(x),\ j,k \in \{0,1\}$. The limit processes for these estimators become

$$\sqrt{n}\left(\widehat{F}^k_{Y_j}(y) - F^k_{Y_j}(y)\right) \Rightarrow \sqrt{\lambda_j} Z^k_j(y) + \sqrt{\lambda_k} \int_{\mathcal{X}} F_{Y_j}(y|x) dB_{X_k}(x), \quad j,k \in \{0,1\},$$

where the familiar first component arises from the estimation of the conditional distribution and the second comes from the estimation of the distributions of the covariates. In Appendix D we discuss further details.

4. Labor Market Institutions and the Distribution of Wages

The empirical application in this section draws its motivation from the influential article by DiNardo, Fortin, and Lemieux (1996, DFL hereafter), which studied the effects of institutional and labor market factors on the evolution of the U.S. wage distribution between 1979 and 1988. The goal of our empirical application is to complete and complement DFL's analysis by using a wider range of techniques, including quantile regression and distribution regression, and to provide confidence intervals for scalar-valued effects as well as function-valued effects of the institutional and labor market factors, such as quantile, distribution, and Lorenz policy effects.

We use the same dataset 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 for men and women separately.

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 also declined from 32 percent to 21 percent for men and from 17 percent to 13 percent for women in our sample, and the composition of the labor force, whose education levels and other characteristics have also 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. The effect (1) measures changes in the marginal distribution of wages that occur due to a change in the minimum wage; the effects (2) and (3) measure changes in the marginal distribution of wages that occur due to a change in the distribution of a particular factor, having fixed the distribution of other factors at some constant level; the effect (4) measures changes in the marginal distribution of wages that occur due to a change in the wage structure, or conditional distribution of wages given worker characteristics.

Next we formally define these four effects as differences between appropriately chosen counterfactual distribution functions. Let $F^{U_r,Z_v}_{Y_t,m_s}$ denote the counterfactual marginal distribution function of log-wages Y when the wage structure is as in year t, the minimum wage, m, is as the level observed for year s, the distribution of union status, U, is as the distribution observed in year r, and the distribution of other worker characteristics, Z, is as the distribution observed in year v. We identify and estimate such counterfactual distributions using the procedures described below. Given these counterfactual distributions, we can decompose the observed total change in the distribution of wages between 1979 and 1988 into the sum of four effects:

$$\begin{aligned} F^{U_{88},Z_{88}}_{Y_{88},m_{88}} - F^{U_{79},Z_{79}}_{Y_{79},m_{79}} &= \underset{(1)}{[F^{U_{88},Z_{88}}_{Y_{88},m_{88}} - F^{U_{88},Z_{88}}_{Y_{88},m_{79}}]} + \underset{(2)}{[F^{U_{88},Z_{88}}_{Y_{88},m_{79}} - F^{U_{79},Z_{88}}_{Y_{88},m_{79}}]} \\ &+ \underset{(3)}{[F^{U_{79},Z_{88}}_{Y_{88},m_{79}} - F^{U_{79},Z_{79}}_{Y_{88},m_{79}}]} + \underset{(4)}{[F^{U_{79},Z_{79}}_{Y_{88},m_{79}} - F^{U_{79},Z_{79}}_{Y_{79},m_{79}}]}. \end{aligned} \tag{4.1}$$

The first component is the effect of the change in the minimum wage, the second is the effect of de-unionization, the third is the effect of changes in worker characteristics, and the fourth is the price effect. As stated above, we see that the effects (2) and (3) measure changes in the marginal distribution of wages that occur due to a change in the distribution

of a particular factor, having fixed the distribution of other factors at some constant level. The effect (4) captures changes in the wage structure or conditional distribution of wages given observed characteristics; in particular, it captures the effect of changes in the market returns to workers' characteristics, including education and experience. Finally, we discuss the interpretation of the minimum wage effect (1) in detail below.

The decomposition (4.1) is the distribution version of the Oaxaca-Blinder decomposition for the mean. We obtain similar decompositions for other functionals $\phi(F^{U_r,Z_v}_{Y_t,m_s})$ of interest, such as marginal quantiles and Lorenz curves, by making an appropriate substitution in equation (4.1) :

$$\begin{aligned}\phi(F^{U_{88},Z_{88}}_{Y_{88},m_{88}}) - \phi(F^{U_{79},Z_{79}}_{Y_{79},m_{79}}) &= \underset{(1)}{[\phi(F^{U_{88},Z_{88}}_{Y_{88},m_{88}}) - \phi(F^{U_{88},Z_{88}}_{Y_{88},m_{79}})]} + \underset{(2)}{[\phi(F^{U_{88},Z_{88}}_{Y_{88},m_{79}}) - \phi(F^{U_{79},Z_{88}}_{Y_{88},m_{79}})]} \\ &+ \underset{(3)}{[\phi(F^{U_{79},Z_{88}}_{Y_{88},m_{79}}) - \phi(F^{U_{79},Z_{79}}_{Y_{88},m_{79}})]} + \underset{(4)}{[\phi(F^{U_{79},Z_{79}}_{Y_{88},m_{79}}) - \phi(F^{U_{79},Z_{79}}_{Y_{79},m_{79}})]}.\end{aligned} \tag{4.2}$$

In constructing the decompositions (4.1) and (4.2), we follow the same sequential order as in DFL.[9] Also, like DFL, we follow a partial equilibrium approach, but, unlike DFL, we do not incorporate supply and demand factors in our analysis because they do not fit well in our framework.

We next describe how to identify and estimate the various counterfactual distributions appearing in (4.1). The first counterfactual distribution we need is $F^{U_{88},Z_{88}}_{Y_{88},m_{79}}$, the distribution of wages that we would observe in 1988 if the real minimum wage were as high as in 1979. Identifying this quantity requires additional assumptions.[10] 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. 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. Under the first strategy, DFL show that

$$F_{Y_{88},m_{79}}(y|u,z) = \begin{cases} F_{Y_{79},m_{79}}(y|u,z)\,\frac{F_{Y_{88},m_{88}}(m_{79}|u,z)}{F_{Y_{79},m_{79}}(m_{79}|u,z)}, & \text{if } y < m_{79}; \\ F_{Y_{88},m_{88}}(y|u,z), & \text{if } y \geq m_{79}; \end{cases} \tag{4.3}$$

[9]The choice of sequential order matters and can affect the relative importance of the four effects. We report some results for the reverse sequential order in the Appendix.

[10]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.

where $F_{Y_t,m_s}(y|u,z)$ denotes the conditional distribution of wages at year t given worker characteristics when the level of the minimum wage is as in year s. Under the second strategy, we have that

$$F_{Y_{88},m_{79}}(y|u,z) = \begin{cases} 0, & \text{if } y < m_{79}; \\ F_{Y_{88},m_{88}}(y|u,z), & \text{if } y \geq m_{79}. \end{cases} \tag{4.4}$$

Given either (4.3) or (4.4) we identify the counterfactual distribution of wages using the representation:

$$F^{U_{88},Z_{88}}_{Y_{88},m_{79}}(y) = \int F_{Y_{88},m_{79}}(y|u,z) dF_{UZ_{88}}(u,z), \tag{4.5}$$

where F_{UZ_t} is the joint distribution of worker characteristics and union status in year t. We can then estimate this distribution using the plug-in principle. In particular, we estimate the conditional distribution in expressions (4.3) and (4.4) using one of the regression methods described below, and the distribution function $F_{UZ_{88}}$ using its empirical analog.

The other counterfactual marginal distributions we need are

$$F^{U_{79},Z_{88}}_{Y_{88},m_{79}}(y) = \int\int F_{Y_{88},m_{79}}(y|u,z)\, dF_{U_{79}}(u|z) dF_{Z_{88}}(z) \tag{4.6}$$

and

$$F^{U_{79},Z_{79}}_{Y_{88},m_{79}}(y) = \int F_{Y_{88},m_{79}}(y|u,z)\, dF_{UZ_{79}}(u,z). \tag{4.7}$$

Given either of our assumptions on the minimum wage 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 $F_{Y_{88},m_{79}}(y|u,z)$ using one of the regression methods described below, the conditional distribution $F_{U_{79}}(u|z), u \in \{0,1\}$, using logistic regression, and $F_{Z_{88}}(z)$ and $F_{UZ_{79}}$ using the empirical distributions.

Formulas (4.5)–(4.7) giving the expressions for the counterfactual distributions reflect the assumptions that give the counterfactual distributions a formal causal interpretation. Indeed, we assume in (4.6) and (4.7) that we can fix the relevant conditional distributions and change only the marginal distributions of the relevant covariates. In (4.5), we also specify how the conditional distribution of wages changes with the level of the minimum wage. Note that we directly observe the marginal distributions appearing on the left side of the decomposition (4.1) and estimate them using the plug-in principle.

To estimate the conditional distributions of wages we consider three different regression methods: classical regression, linear quantile regression, and distribution regression with a logit link. The classical regression, despite its wide use in the literature, is not appropriate in this application due to substantial conditional heteroscedasticity in log wages

(Lemieux, 2006, and Angrist, Chernozhukov, and Fernandez-Val, 2006). The linear quantile regression is more flexible, but it also has shortcomings in this application. First, there is a considerable amount of rounding, especially at the level of the minimum wage, which makes the wage variable highly discrete. Second, 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. The distribution regression approach does not suffer from these problems, and we therefore employ it to generate the main empirical results. In order to check the robustness of our empirical results, we also employ the censoring approach described above. We set the wages below the minimum wage to the value of the minimum wage and then apply censored quantile and distribution regressions to the resulting data. In what follows, we first present the empirical results obtained using distribution regression, and then briefly compare them with the results obtained using censored quantile regression and censored distribution regression.

We present our empirical results in Tables 1–3 and Figures 1–9. In Figure 1, we compare the empirical distributions of wages in 1979 and 1988. In Table 1, we report the estimation and inference results for the decomposition (4.2) of the changes in various measures of wage dispersion between 1979 and 1988 estimated using distribution regressions.[11] Figures 2–7 refine these results by presenting estimates and 95% simultaneous confidence intervals for several major functionals of interest, including the effects on entire quantile functions, distribution functions, and Lorenz curves. 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. In Tables 2–3 and Figures 8–9, we present the estimates of the same effects as in Table 1 and Figures 2–3 estimated using various alternative methods, such as censored quantile regression and censored distribution regression. Overall, we find that our estimates, confidence intervals, and robustness checks all reinforce the findings of DFL, giving them a rigorous econometric foundation. Indeed, we provide standard errors and confidence intervals, without which we would not be able to assess the statistical significance of the results. Moreover, we validate the results with a wide array of estimation methods. In what follows below, we discuss each of our results in more detail.

[11]The estimation results parallel the results presented in DFL. Table A1 in the Appendix gives the results for the decomposition in reverse order.

In Figure 1, we present estimates and uniform confidence intervals for the marginal distributions of wages in 1979 and 1988. We see 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. In Figures 2–3, we present estimates and uniform confidence intervals for the total change in the marginal quantile function of wages and the four effects that form a decomposition of this total change.[12] We report the marginal quantile functions in 1979 and 1988 in the top left panels of Figures 2 and 3. In Figures 4–7, we plot analogous results for the decomposition of the total change in marginal distribution functions and Lorenz curves.

From Figures 2 and 3, 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 for men. 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), and illustrates the difference between conditional and unconditional effects.[13] In general, interpreting the unconditional effect of changes in the distribution of a covariate requires some care, because the covariate may change only over certain parts of its support. For example, de-unionization cannot affect those who were not unionized at the beginning of the period, which is 70 percent of the workers; and in our data, the unionization 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 since only 9 percent of the workers are affected.

From Figures 2 and 3, 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

[12]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 Tables 1–3 and A1 have not been smoothed.

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

has been recently stressed by Lemieux (2006) and Autor, Katz and Kearney (2008). The composition effect is realized through at least two channels. The first channel operates through between-group inequality. In our case, higher 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, higher 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:

$$Var[Y] = E[\beta(U)]'Var[X]E[\beta(U)] + trace\{E[XX']Var[\beta(U)]\}. \tag{4.8}$$

The first channel corresponds to changes in the first term of (4.8) 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 (4.8) operating through the interaction of between group inequality $E[XX']$ and within group inequality $Var[\beta(U)]$.

In Figures 2 and 3, 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 Tables 2–3 and Figures 8–9, we present several interesting robustness checks. As we mentioned above, the assumptions about the minimum wage are particularly delicate, since the mechanism that generates wages strictly below this level is not clear; it could be measurement error, non-coverage, or non-compliance with the law. To check the robustness of the results to the DFL assumptions about the minimum wage and to our semi-parametric

model of the conditional distribution, we re-estimate the decomposition using censored linear quantile regression and censored distribution regression with a logit link, using the wage data censored below the minimum wage. For censored quantile regression, we use Powell's (1986) censored quantile regression estimated using Chernozhukov and Hong's (2002) algorithm. For censored distribution regression, we simply censor to zero the distribution regression estimates of the conditional distributions below the minimum wage and recompute the functionals of interest. Overall, we find the results are very similar for the quantile and distribution regressions, and they are not very sensitive to the censoring.[14]

5. Conclusion

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 estimator of the conditional distribution or conditional quantile function. This condition holds for most important semiparametric estimators of conditional distribution and quantile functions, such as classical, quantile, duration, and distribution regressions.

† *Massachusetts Institute of Technology, Department of Economics & Operations Research Center; and University College London, CEMMAP. E-mail: vchern@mit.edu.* *Research support from the Castle Krob Chair, National Science Foundation, the Sloan Foundation, and CEMMAP is gratefully acknowledged.*

§ *Boston University, Department of Economics. E-mail: ivanf@bu.edu.* *Research support from the National Science Foundation is gratefully acknowledged.*

‡ *Brown University, Department of Economics. E-mail: Blaise_Melly@brown.edu.*

Appendix

This Appendix contains proofs and additional results. Section A collects preliminary lemmas on the functional delta method and derives the functional delta method for any simulation method, extending its applicability beyond the bootstrap. Section B collects the proofs for the results in the main text of the paper. Section C gives limit distribution theory for policy effects estimators. Section D presents additional results for the case where the covariate distributions are estimated. These results complement the results in

[14]We have additional results on quantile, distribution and Lorenz effects for the censored estimates; these are available on request from the authors. We do not report them here to save space.

the main text. Section E derives limit theory, including Hadamard differentiability, for Z-processes and Section F applies this theory to the principal estimators of conditional distribution and quantile functions. These results establish the validity of bootstrap and other resampling schemes for the entire quantile regression process, the entire distribution regression process, and related processes arising in the estimation of various conditional quantile and distribution functions. These results may be of a substantial independent interest.

Appendix A. Functional Delta Method, Bootstrap, and Other Methods

This section collects preliminary lemmas on the functional delta method and derives the functional delta method for any simulation method, extending its applicability beyond the bootstrap.

A.1. **Some definitions and auxiliary results.**

We begin by quickly recalling from van der Vaart and Wellner (1996) the details of the functional delta method.

Definition 1 (Hadamard-differentiability)**.** *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.

This notion works well together with the continuous mapping theorem.

Lemma 2 (Extended continuous mapping theorem)**.** *Let $\mathbb{D}_n \subset \mathbb{D}$ be arbitrary subsets and $g_n : \mathbb{D}_n \mapsto \mathbb{E}$ be arbitrary maps ($n \geq 0$), such that for every sequence $x_n \in \mathbb{D}_n$: if $x_{n'} \to x \in \mathbb{D}_0$ along a subsequence, then $g_{n'}(x_{n'}) \to g_0(x)$. Then, for arbitrary maps $X_n : \Omega_n \mapsto \mathbb{D}_n$ and every random element X with values in $\mathbb{D}_0$ such that $g_0(X)$ is a random element in $\mathbb{E}$:*

(i) *If $X_n \Rightarrow X$, then $g_n(X_n) \Rightarrow g_0(X)$;*
(ii) *If $X_n \to_p X$, then $g_n(X_n) \to_p g_0(X)$.*

The combination of the previous definition and lemma is known as the functional delta method.

Lemma 3 (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 $r_n(X_n - \theta) \Rightarrow X$ 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 $r_n(\phi(X_n) - \phi(\theta)) \Rightarrow \phi'_\theta(X)$. 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.

Lemma 4 (Chain rule)**.** *If $\phi : \mathbb{D}_\phi \subset \mathbb{D} \mapsto \mathbb{E}_\psi$ is Hadamard-differentiable at $\theta \in \mathbb{D}_\phi$ tangentially to $\mathbb{D}_0$ and $\psi : \mathbb{E}_\psi \mapsto \mathbb{F}$ is Hadamard-differentiable at $\phi(\theta)$ tangentially to $\phi'(\mathbb{D}_0)$, then $\psi \circ \phi : \mathbb{D}_\phi \mapsto \mathbb{F}$ is Hadamard-differentiable at θ tangentially to $\mathbb{D}_0$ with derivative $\psi'_{\phi(\theta)} \circ \phi'_\theta$.*

Another technical result to be used in the sequel is concerns the equivalence of continuous and uniform convergence.

Lemma 5 (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 $f_n(x_n) \to f(x)$.*

Proof of Lemmas 2–4: See van der Vaart and Wellner (1996) Chap. 1.11 and 3.9. □

Proof of Lemma 5: See, for example, Resnick (1987), page 2. □

A.2. **Functional delta-method for bootstrap and other simulation methods.** Let $\mathcal{F}_n = (W_1, ..., W_n)$ denote the data. Consider sequences of random elements $V_n = V_n(\mathcal{F}_n)$, the original empirical process. In a normed space $\mathbb{D}$, the sequence $\sqrt{n}(V_n - V)$ converges unconditionally to the process $\mathbb{G}$. Let the sequence of random elements

$$\widehat{V}_n = V_n + G_n/\sqrt{m} \tag{A.1}$$

where $m = m(n)$ is a possibly random sequence such that $m/m_0 \to_P 1$ for some sequence of constants $m_0 \to \infty$ such that $m_0/n \to c \geq 0$,[15] and the "draw" G_n is produced by

[15]The random scaling is needed to cover wild bootstrap, for example.

bootstrap, simulation, or any other consistent method that guarantees that the sequence G_n converges conditionally given $\mathcal{F}_n$ in distribution to a tight random element $\mathbb{G}$,

$$\sup_{h\in \mathrm{BL}_1(\mathbb{D})} \left| E_{|\mathcal{F}_n} h\left(G_n\right)^* - Eh(\mathbb{G})\right| \to 0, \tag{A.2}$$

in outer probability, where $\mathrm{BL}_1(\mathbb{D})$ denotes the space of function with Lipschitz norm at most 1 and $E_{|\mathcal{F}_n}$ denotes the conditional expectation given the data. In the definition, we can take $\mathbb{G}$ to be independent of $\mathcal{F}_n$.

Given a map $\phi : \mathbb{D}_\phi \subset \mathbb{D} \mapsto \mathbb{E}$, we wish to show that

$$\sup_{h\in \mathrm{BL}_1(\mathbb{E})} \left| E_{|\mathcal{F}_n} h\left(\sqrt{m}(\phi(\widehat{V}_n) - \phi(V_n))\right)^* - Eh(\phi'_V(\mathbb{G}))\right| \to 0, \tag{A.3}$$

in outer probability.

Lemma 6 (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 $\widehat{V}_n$ be maps as indicated previously with values in $\mathbb{D}_\phi$ such that $\sqrt{n}(V_n - V) \Rightarrow \mathbb{G}$ and (A.2) holds in outer probability, where $\mathbb{G}$ is separable and takes its values in $\mathbb{D}_0$. Then (A.3) holds in outer probability.*

Proof of Lemma 6: The proof generalizes the functional delta-method for empirical bootstrap in Theorem 3.9.11 of van der Vaart and Wellner (1996) to exchangeable bootstrap. This expands the applicability of delta-method to a wide variety of resampling and simulation schemes that are special cases of exchangeable bootstrap, including empirical bootstrap, Bayesian bootstrap, wild bootstrap, k out of n bootstrap, and subsampling bootstrap (see next section for details).

Without loss of generality, assume that the derivative $\phi'_V : \mathbb{D} \mapsto \mathbb{E}$ is defined and continuous on the whole space. Otherwise, replace $\mathbb{E}$ by its second dual $\mathbb{E}^{**}$ and the derivative by an extension $\phi'_V : \mathbb{D} \mapsto \mathbb{E}^{**}$. For every $h \in \mathrm{BL}_1(\mathbb{E})$, the function $h \circ \phi'_V$ is contained in $\mathrm{BL}_{\|\phi'_V\|}(\mathbb{D})$. Thus (A.2) implies $\sup_{h\in \mathrm{BL}_1(\mathbb{E})} \left| E_{|\mathcal{F}_n} h\left(\phi'_V(G_n)\right)^* - Eh(\phi'_V(\mathbb{G}))\right| \to 0$, in outer probability. Next

$$\begin{aligned} &\sup_{h\in \mathrm{BL}_1(\mathbb{E})} \left| E_{|\mathcal{F}_n} h\left(\sqrt{m}\left(\phi(\widehat{V}_n) - \phi(V_n)\right)\right)^* - E_{|\mathcal{F}_n} h\left(\phi'_V\left(G_n\right)\right)_*\right| \\ &\leq \varepsilon + 2\mathrm{P}_{|\mathcal{F}_n}\left(\left\|\sqrt{m}\left(\phi(\widehat{V}_n) - \phi(V_n)\right) - \phi'_V\left(\sqrt{m}(\widehat{V}_n - V_n)\right)\right\|^* > \varepsilon\right). \end{aligned} \tag{A.4}$$

The theorem is proved once it has been shown that the conditional probability on the right converges to zero in outer probability.

Both sequences $\sqrt{m}(V_n - V)$ and $G_n = \sqrt{m}(\widehat{V}_n - V)$ converge (unconditionally) in distribution to separable random elements that concentrate on the space $\mathbb{D}_0$. The first

sequence converges by assumption and Slutsky's theorem when $m/m_0 \to_P 1$ and $m_0/n \to c > 0$ and converges to zero when $m_0/n \to 0$ by assumption and Slutsky's theorem. The second sequence converges, by noting that

$$\sqrt{m}(\widehat{V}_n - V) = \sqrt{m}(\widehat{V}_n - V_n) + \underbrace{\sqrt{m}(V_n - V)}_{t_n}$$

and that $E|E_{|\mathcal{F}_n}h(\sqrt{m}(\widehat{V}_n - V_n)^* + t_n) - E_{|\mathcal{F}_n}h(\mathbb{G} + t_n)| \leq \sup_{h\in \mathrm{BL}_1(\mathbb{D}_n)} E|E_{|\mathcal{F}_n}h(\sqrt{m}(\widehat{V}_n - V_n))^* - E_{|\mathcal{F}_n}h(\mathbb{G})| = \sup_{h\in \mathrm{BL}_1(\mathbb{D}_n)} E|E_{|\mathcal{F}_n}h((G_n)^* - E_{|\mathcal{F}_n}h(\mathbb{G})|$ which converges to zero by (A.2), and by $E_{|\mathcal{F}_n}h(\mathbb{G}) = Eh(\mathbb{G})$ due to independence of $\mathbb{G}$ from $\mathcal{F}_n$.

By Lemma 3,

$$\begin{aligned} \sqrt{m}\left(\phi(\widehat{V}_n) - \phi(V)\right) &= \phi'_V\left(\sqrt{m}(\widehat{V}_n - V)\right) + o_P^*(1), \\ \sqrt{m}\left(\phi(V_n) - \phi(V)\right) &= \phi'_V\left(\sqrt{m}(V_n - V)\right) + o_P^*(1). \end{aligned} \tag{A.5}$$

Subtract these equations to conclude that the sequence $\sqrt{m}(\phi(\widehat{V}_n) - \phi(V_n)) - \phi'_V(\sqrt{m}(\widehat{V}_n - V_n))$ converges unconditionally to zero in outer probability. Thus, the conditional probability on the right in (A.4) converges to zero in outer mean. □

A.3. **Exchangeable Bootstrap.** Let $(W_1, ..., W_n)$ denote the i.i.d. data. Next we define the collection of exchangeable bootstrap methods that we can employ for inference. For each n, let $(e_{n1}, ..., e_{nn})$ be an exchangeable, nonnegative random vector. Exchangeable bootstrap uses the components of this vector as random sampling weights in place of constant weights $(1, ..., 1)$. A simple way to think of exchangeable bootstrap is as sampling each variable W_i the number of times equal to e_{ni}, albeit without requiring e_{ni} to be integer-valued. Given an empirical process $V_n(f) = \frac{1}{n}\sum_{i=1}^n f(X_i)$, we define an exchangeable bootstrap draw of this process as

$$\widehat{V}_n(f) := V_n(f) + G_n(f)/\sqrt{m}, \quad m = n\bar{e}_n^2, \quad G_n(f) := \frac{1}{\sqrt{n}}\sum_{i=1}^n (e_{ni} - \bar{e}_n)f(W_i),$$

where $\bar{e}_n = \sum_{i=1}^n e_{ni}/n$. This insures that each draw of $\widehat{V}_n$ assigns nonnegative weights to each observation, which is important in applications of bootstrap to extremum estimators to preserve convexity of criterion functions. We assume that, for some $\varepsilon > 0$

$$\sup_n E[e_{n1}^{2+\varepsilon}] < \infty, \quad n^{-1}\sum_{i=1}^n (e_{ni} - \bar{e}_n)^2 \to_P 1, \quad \bar{e}_n^2 \to_P c \geq 0, \tag{A.6}$$

where the first two conditions are standard, see Van der Vaart and Wellner (1996), and the last one is needed to apply the previous lemma. Let us consider the following special cases: (1) The standard empirical bootstrap corresponds to the case where $(e_{n1}, ..., e_{nn})$

is a multinomial vector with parameters n and probabilities $(1/n, ..., 1/n)$, so that $\bar{e}_n = 1$ and $m = n$. (2) The Bayesian bootstrap corresponds to the case where $U_1, ..., U_n$ are i.i.d. nonnegative random variables, e.g. unit exponential, with $E|U_1^{2+\varepsilon}] < \infty$ for some $\varepsilon > 0$, and $e_{ni} = U_i/\bar{U}_n$, so that $\bar{e}_n = 1$ and $m = n$. (3) The wild bootstrap corresponds to the case where $e_{n1}, ..., e_{nn}$ are i.i.d. vectors with $E[e_{n1}^{2+\varepsilon}] < \infty$ for some $\varepsilon > 0$, and $Var[e_{n1}] = 1$, so that $m/n = \bar{e}_n^2 \to_P Ee_{n1}^2 \geq 0$ and $m_0 = nEe_{n1}^2 \to \infty$. (4) The k out of n bootstrap resamples $k < n$ observations from $W_1, ..., W_n$ with replacement. This corresponds to letting $(e_{n1}, ..., e_{nn})$ be equal to $\sqrt{n/k}$ times multinomial vectors with parameters k and probabilities $(1/n, ..., 1/n)$. The condition (A.6) on the weights holds if $k \to \infty$, so that $\bar{e}_n^2 = k/n \to c \geq 0$ and $m = k \to \infty$. (5) The subsampling bootstrap corresponds to resampling $k < n$ observations from $W_1, ..., W_n$ without replacement. This corresponds to letting $(e_{n1}, ..., e_{nn})$ be a row of k times the number $n(n-k)^{-1/2}k^{-1/2}$ and $n - k$ times the number 0, ordered at random, independent of the W_i's. The condition (A.6) on the weights holds if both $k \to \infty$ and $n - k \to \infty$. In this case $\bar{e}_n^2 = k/(n-k) \to c \geq 0$ and $m = nk/(n-k) \to \infty$.

As a consequence of Lemma 6, we obtain the following result, which might be of independent interest.

Lemma 7 (Functional delta method for exchangeable bootstrap)**.** *The exchangeable bootstrap method described above satisfies condition (A.2), and therefore the conclusions of Lemma 6 about validity of the functional delta method apply to this method.*

Proof of Lemma 7: By Lemma 6, we only need to verify condition (A.2), which follows by Theorem 3.6.13 of Van der Vaart and Wellner (1996). □

Appendix B. Inference Theory for Counterfactual Estimators (Proofs)

This section collects the proofs for the results in the main text of the paper.

B.1. **Notation.** Define $Y_x := Q_Y(U|x)$, where $U \sim \text{Uniform}(\mathcal{U})$ with $\mathcal{U} = (0, 1)$. Denote by $\mathcal{Y}_x$ the support of Y_x, $\mathcal{YX} := \{(y, x) : y \in \mathcal{Y}_x, x \in \mathcal{X}\}$, and $\mathcal{UX} := \mathcal{U} \times \mathcal{X}$. We assume throughout that $\mathcal{Y}_x \subset \mathcal{Y}$, which is a compact subset of $\mathbb{R}$, and that $x \in \mathcal{X}$, a compact subset of $\mathbb{R}^p$. In what follows, $\ell^\infty(\mathcal{UX})$ denotes the set of bounded and measurable functions $h : \mathcal{UX} \mapsto \mathbb{R}$, and $C(\mathcal{UX})$ denotes the set of continuous functions mapping $h : \mathcal{UX} \mapsto \mathbb{R}$.

B.2. **Uniform Hadamard differentiability of conditional distribution functions with respect to the conditional quantile functions.** The following lemma establishes

the Hadamard differentiability of the conditional distribution function with respect to the conditional quantile function. We use this result to prove Lemma 1 in the main text and to derive the limit distribution for the policy estimators based on conditional quantile models. We drop the dependence on the group index to simplify the notation.

Lemma 8 (Hadamard derivative of $F_Y(y|x)$ with respect to $Q_Y(u|x)$)**.** *Define* $F_Y(y|x, h_t) := \int_0^1 1\{Q_Y(u|x) + th_t(u|x) \leq y\}du$. *Under condition C, as* $t \searrow 0$,

$$D_{h_t}(y|x,t) = \frac{F_Y(y|x,h_t) - F_Y(y|x)}{t} \to D_h(y|x) := -f_Y(y|x)h(F_Y(y|x)|x).$$

The convergence holds uniformly in any compact subset of $\mathcal{YX} := \{(y,x) : y \in \mathcal{Y}_x, x \in \mathcal{X}\}$, *for every* $\|h_t - h\|_\infty \to 0$, *where* $h_t \in \ell^\infty(\mathcal{UX})$, *and* $h \in C(\mathcal{UX})$.

Proof of Lemma 8: We have that for any $\delta > 0$, there exists $\epsilon > 0$ such that for $u \in B_\epsilon(F_Y(y|x))$ and for small enough $t \geq 0$

$$1\{Q_Y(u|x) + th_t(u|x) \leq y\} \leq 1\{Q_Y(u|x) + t(h(F_Y(y|x)|x) - \delta) \leq y\};$$

whereas for all $u \notin B_\epsilon(F_Y(y|x))$,

$$1\{Q_Y(u|x) + th_t(u|x) \leq y\} = 1\{Q_Y(u|x) \leq y\}.$$

Therefore, for small enough $t \geq 0$

$$\begin{aligned} &\frac{\int_0^1 1\{Q_Y(u|x) + th_t(u|x) \leq y\}du - \int_0^1 1\{Q_Y(u|x) \leq y\}du}{t} \\ \leq &\int_{B_\epsilon(F_Y(y|x))} \frac{1\{Q_Y(u|x) + t(h(F_Y(y|x)|x) - \delta) \leq y\} - 1\{Q_Y(u|x) \leq y\}}{t} du, \end{aligned} \tag{B.1}$$

which by the change of variable $\tilde{y} = Q_Y(u|x)$ is equal to

$$\frac{1}{t}\int_{J \cap [y, y - t(h(F_Y(y|x)|x) - \delta)]} f_Y(\tilde{y}|x)d\tilde{y},$$

where J is the image of $B_\epsilon(F_Y(y|x))$ under $u \mapsto Q_Y(\cdot|x)$. The change of variable is possible because $Q_Y(\cdot|x)$ is one-to-one between $B_\epsilon(F_Y(y|x))$ and J.

Fixing $\epsilon > 0$, for $t \searrow 0$, we have that $J \cap [y, y - t(h(F_Y(y|x)|x) - \delta)] = [y, y - t(h(F_Y(y|x)|x) - \delta)]$, and $f_Y(\tilde{y}|x) \to f_Y(y|x)$ as $F_Y(\tilde{y}|x) \to F_Y(y|x)$. Therefore, the right hand term in (B.1) is no greater than

$$-f_Y(y|x)\left(h(F_Y(y|x)|x) - \delta\right) + o\left(1\right).$$

Similarly $-f_Y(y|x)\left(h(F_Y(y|x)|x) + \delta\right) + o\left(1\right)$ bounds (B.1) from below. Since $\delta > 0$ can be made arbitrarily small, the result follows.

To show that the result holds uniformly in $(y, x) \in K$, a compact subset of $\mathcal{YX}$, we use Lemma 5. Take a sequence of (y_t, x_t) in K that converges to $(y, x) \in K$, then the preceding argument applies to this sequence, since the function $(y, x) \mapsto -f_Y(y|x)h(F_Y(y|x)|x)$ is uniformly continuous on K. This result follows by the assumed continuity of $h(u|x)$, $F_Y(y|x)$ and $f_Y(y|x)$ in both arguments, and the compactness of K. □

B.3. **Proof of Lemma 1.** This result follows by the Hadamard differentiability of the conditional distribution function with respect to the conditional quantile function in Lemma 8, Condition Q, and the functional delta method in Lemma 3. □

B.4. **Proof of Theorem 1.** The joint uniform convergence result follows from Condition D by the extended continuous mapping theorem in Lemma 2, since the integral is a continuous operator. Gaussianity of the limit process follows from linearity of the integral. □

B.5. **Proof of Theorem 2.** The joint uniform convergence result and Gaussianity of the limit process follow from Theorem 1 by the functional delta method in Lemma 3 , since the quantile operator is Hadamard differentiable (see, e.g., Doss and Gill, 1992). □

B.6. **Proof of Corollary 1.** This result follows from Theorem 2 by the extended continuous mapping theorem in Lemma 2. □

B.7. **Proof of Corollary 2.** This result follows from Theorem 1 by the extended continuous mapping theorem in Lemma 2. □

B.8. **Proof of Corollary 3.** This result follows from Theorem 1 by the functional delta method in Lemma 3 and the chain rule for Hadamard differentiable functionals in Lemma 4. □

B.9. **Proof of Theorem 3.** This result follows from the functional delta method for the bootstrap and other simulation methods in Lemma 6. □

Appendix C. Limit distribution for the estimators of the effects

For policy interventions that can be implemented either as a known transformation of the covariate, $X_1 = g(X_0)$, or as a change in the conditional distribution of Y given X, we can also identify and estimate the distribution of the effect of the policy, $\Delta_j^k = Y_j^k - Y_0^0$, $j, k \in \{0, 1\}$, under Condition RP stated in the main text. The following

results provide estimators for the distribution and quantile functions of the effects and limit distribution theory for them. Let $\mathcal{D} = \{\delta \in \mathbb{R} : \delta = y - \tilde{y}, y \in \mathcal{Y}, \tilde{y} \in \mathcal{Y}\}$.

Lemma 9 (Limit distribution for estimators of conditional distribution and quantile functions)**.** *Let $\widehat{Q}_{\Delta_0}(u|x) = \widehat{Q}_{Y_0}(u|g(x)) - \widehat{Q}_{Y_0}(u|x)$ and $\widehat{Q}_{\Delta_1}(u|x) = \widehat{Q}_{Y_1}(u|x) - \widehat{Q}_{Y_0}(u|x)$ be estimators of the conditional quantile function of the effect $Q_{\Delta_j}(u|x), j \in \{0,1\}$.*[16] *Under the conditions C, Q, and RP, we have:*

$$\sqrt{n}\left(\widehat{Q}_{\Delta_j}(u|x) - Q_{\Delta_j}(u|x)\right) \Rightarrow V_{\Delta_j}(u,x),\ j \in \{0,1\},$$

in $\ell^\infty((0,1) \times \mathcal{X})$, where $V_{\Delta_0}(u,x) := \sqrt{\lambda_0}[V_0(u,g(x)) - V_0(u,x)]$ and $V_{\Delta_1}(u,x) := \sqrt{\lambda_1}V_1(u,x) - \sqrt{\lambda_0}V_0(u,x)$. The Gaussian processes $(u,x) \mapsto V_{\Delta_j}(u,x),\ j \in \{0,1\}$, have zero mean and covariance function $\Omega_{V_{jr}}(u,x,\tilde{u},\tilde{x}) := E[V_{\Delta_j}(u,x)V_{\Delta_r}(\tilde{u},\tilde{x})]$, for $j,r \in \{0,1\}$.

Let $\widehat{F}_{\Delta_j}(\delta|x) = \int_0^1 1\{\widehat{Q}_{\Delta_j}(u|x) \leq \delta\}du$ be an estimator of the conditional distribution of the effects $F_{\Delta_j}(\delta|x)$, for $j \in \{0,1\}$. Under the conditions C, Q, and RP, we have:

$$\sqrt{n}\left(\widehat{F}_{\Delta_j}(\delta|x) - F_{\Delta_j}(\delta|x)\right) \Rightarrow -f_{\Delta_j}(\delta|x)V_{\Delta_j}(F_{\Delta_j}(\delta|x),x) =: Z_{\Delta_j}(\delta,x),\ j \in \{0,1\},$$

in $\ell^\infty(\mathcal{D} \times \mathcal{X})$, and $(\delta,x) \mapsto Z_{\Delta_j}(\delta,x), j \in \{0,1\}$, have zero mean and covariance function $\Omega_{Z_{jr}}(\delta,x,\tilde{\delta},\tilde{x}) := E[Z_{\Delta_j}(\delta,x)Z_{\Delta_r}(\tilde{\delta},\tilde{x})]$, for $j,r \in \{0,1\}$. The conditional density of the effect, $f_{\Delta_j}(\delta|x)$, is assumed to be bounded above and away from zero.[17]

Proof of Lemma 9. The uniform convergence result for the conditional quantile processes $\sqrt{n}(\widehat{Q}_{\Delta_j}(u|x) - Q_{\Delta_j}(u|x)), j \in \{0,1\}$, follows from Conditions Q and RP by the extended continuous mapping theorem in Lemma 2. Uniform convergence of the conditional distribution processes $\sqrt{n}(\widehat{F}_{\Delta_j}(\delta|x) - F_{\Delta_j}(\delta|x)), j \in \{0,1\}$, follows from the covergence of the quantile process by the functional delta method in Lemma 3. The Hadamard differentiability of $F_{\Delta_j}(\delta|x)$ with respect to $Q_{\Delta_j}(u|x)$ can be established using the same argument as in the proof of Lemma 8. □

Theorem 4 (Limit distribution for estimators of the marginal distribution and quantile functions)**.** *Under the conditions M, C, Q, and RP, the estimators $\widehat{F}^k_{\Delta_j}(\delta) =$*

[16] In the distribution approach, $\widehat{Q}_{Y_j}(u|x)$ can be obtained by inversion of the estimator of the conditional distribution.

[17] This assumption rules out degenerated distributions for the distribution of effects, such as constant policy effects. These "distributions" can be estimated using standard regression methods.

$\int_{\mathcal{X}} \widehat{F}_{\Delta_j}(\delta|x)dF_{X_k}(x)$ *of the marginal distributions of the effects* $F^k_{\Delta_j}(\delta)$ *jointly converge in law to the following Gaussian processes:*

$$\sqrt{n}\left(\widehat{F}^k_{\Delta_j}(\delta) - F^k_{\Delta_j}(\delta)\right) \Rightarrow \int_{\mathcal{X}} Z_{\Delta_j}(\delta, x)dF_{X_k}(x) =: Z^k_{\Delta_j}(\delta),\ j,k \in \{0,1\},$$

in $\ell^\infty(\mathcal{D})$*, where* $\delta \mapsto Z^k_{\Delta_j}(\delta)$*,* $j,k \in \{0,1\}$*, have zero mean and covariance function* $\Omega^{ks}_{Z_{jr}}(\delta,\tilde{\delta}) := E[Z^k_{\Delta_j}(\delta)Z^s_{\Delta_r}(\tilde{\delta})]$*, for* $j,k,r,s \in \{0,1\}$*.*

Under the conditions M, C, Q, and RP, the estimators $\widehat{Q}^k_{\Delta_j}(u) = \inf\{\delta : \widehat{F}^k_{\Delta_j}(\delta) \geq u\}$ *of the marginal quantile functions of the effects* $Q^k_{\Delta_j}(u)$ *jointly converge in law to the following Gaussian processes:*

$$\sqrt{n}\left(\widehat{Q}^k_{\Delta_j}(u) - Q^k_{\Delta_j}(u)\right) \Rightarrow -Z^k_{\Delta_j}(Q^k_{\Delta_j}(u))/f^k_{\Delta_j}(Q^k_{\Delta_j}(u)) =: V^k_{\Delta_j}(u),\ j,k \in \{0,1\},$$

in $\ell^\infty((0,1))$*, where* $f^k_{\Delta_j}(\delta) = \int_{\mathcal{X}} f_{\Delta_j}(\delta|x)dF_{X_k}(x)$ *and* $u \mapsto V^k_{\Delta_j}(u)$*,* $j \in \{0,1\}$*, have zero mean and variance function* $\Omega^{ks}_{V_{jr}}(u,\tilde{u}) := E[V^k_{\Delta_j}(u)V^s_{\Delta_r}(\tilde{u})]$*, for* $j,k,r,s \in \{0,1\}$*.*

Proof of Theorem 4. The uniform convergence result for the marginal distribution functions follows from the convergence of the conditional processes in Lemma 9 by the extended continuous mapping theorem in Lemma 2, since the integral is a continuous operator. Gaussianity of the limit process follows from linearity of the integral. The uniform convergence result for the quantile function follows from the convergence of the distribution function by the functional delta method in Lemma 3, since the quantile operator is Hadamard differentiable (see, e.g., Doss and Gill, 1992). □

Appendix D. Inference Theory for Counterfactuals Estimators: The Case with Estimated Covariate Distributions

This section presents additional results for the case where the covariate distributions are estimated. These results complement the analysis in the main text.

D.1. **Limit theory, bootstrap, and other simulation methods.** We start by restating Condition D to incorporate the assumptions about the estimators of the covariate distributions.

Condition DC. *(a) Let* $\widehat{Z}_j(y,x) := \sqrt{n}(\widehat{F}_{Y_j}(y|x) - F_{Y_j}(y|x))$ *and* $\widehat{G}_{X_k}(f) := \sqrt{n}\int f d(\widehat{F}_{X_k}(x) - F_{X_k}(x))$*, where* $\widehat{F}_{X_k}$ *are estimated probability measures, for* $j,k \in \{0,1\}$*. These measures must support the P-Donsker property, namely*

$$\left(\widehat{Z}_0, \widehat{Z}_1, \widehat{G}^0_X, \widehat{G}^1_X\right) \Rightarrow \left(\sqrt{\lambda_0}Z_0, \sqrt{\lambda_1}Z_1, \sqrt{\lambda_0}G_{X_0}, \sqrt{\lambda_1}G_{X_1}\right),$$

in the space $\ell^\infty(\mathcal{Y}\times\mathcal{X})\times\ell^\infty(\mathcal{Y}\times\mathcal{X})\times\ell^\infty(\mathcal{F})\times\ell^\infty(\mathcal{F})$*, for each* F_X*-Donsker class* $\mathcal{F}$*, where the right hand side is a zero mean Gaussian process and* λ_j *is the limit of the ratio of the sample size in group* j *to the total sample size* n*, for* $j\in\{0,1\}$*.*

(b) The function class $\{F_{Y_j}(y|X), y\in\mathcal{Y}\}$ *is* F_{X_k}*-Donsker, for* $j,k\in\{0,1\}$*.*

The condition on the estimated measure is weak and is satisfied when $\widehat{F}_{X_j}$ is an empirical measure based on a random sample. Moreover, the condition holds for various smooth empirical measures; in fact, in this case the class of functions $\mathcal{F}$ for which DC(a) holds can be much larger than Glivenko-Cantelli or Donsker (see Radulovic and Wegkamp, 2003, and Gine and Nickl, 2008). Condition DC(b) is also a weak condition that holds for rich classes of functions, see, e.g., van der Vaart (1998).

Theorem 5 (Limit distribution and inference theory for counterfactual marginal distributions)**.** *(1) Under conditions M and DC the estimators* $\widehat{F}^k_{Y_j}(y)=\int_{\mathcal{X}}\hat{F}_{Y_j}(y|x)d\widehat{F}_{X_k}(x)$ *of the marginal distribution functions* $F^k_{Y_j}(y)$ *jointly converge in law to the following Gaussian processes:*

$$\sqrt{n}\left(\widehat{F}^k_{Y_j}(y)-F^k_{Y_j}(y)\right)\Rightarrow\sqrt{\lambda_j}Z^k_j(y)+\sqrt{\lambda_k}G_{X_k}(F_{Y_j}(y|\cdot))=:\tilde{Z}^k_j(y),\quad j,k\in\{0,1\},\ \text{(D.1)}$$

in $\ell^\infty(\mathcal{Y})$*, where* $y\mapsto\tilde{Z}^k_j(y),\ j,k\in\{0,1\}$*, have zero mean and covariance function, for* $j,k,r,s\in\{0,1\}$*,*

$$\tilde{\Sigma}^{ks}_{Z_{jr}}(y,\tilde{y}):=\sqrt{\lambda_j\lambda_r}\Sigma^{ks}_{Z_{jr}}(y,\tilde{y})+\sqrt{\lambda_k\lambda_s}E\left[G_{X_k}(F_{Y_j}(y|\cdot))G_{X_s}(F_{Y_r}(\tilde{y}|\cdot))\right],\qquad\text{(D.2)}$$

where $\Sigma^{ks}_{Z_{jr}}$ *is defined as in (3.4).*

(2) Any bootstrap or other simulation method that consistently estimates the law of the empirical process $(\widehat{Z}_0,\widehat{Z}_1,\widehat{G}_{X_0},\widehat{G}_{X_1})$ *in the space* $\ell^\infty(\mathcal{Y}\times\mathcal{X})\times\ell^\infty(\mathcal{Y}\times\mathcal{X})\times\ell^\infty(\mathcal{F})\times\ell^\infty(\mathcal{F})$*, also consistently estimates the law of the empirical process* $(\tilde{Z}^0_0,\tilde{Z}^1_1,\tilde{Z}^1_0,\tilde{Z}^0_1)$ *in the space* $\ell^\infty(\mathcal{Y})\times\ell^\infty(\mathcal{Y})\times\ell^\infty(\mathcal{Y})\times\ell^\infty(\mathcal{Y})$*.*

Proof of Theorem 5: The first part of the theorem follows by the functional delta method in Lemma 3 and the Hadamard differentiability of the marginal functions demonstrated in Lemma 10 below with $t=1/\sqrt{n}$. The second part of the theorem follows by the functional delta method for the bootstrap and other simulation methods in Lemma 6. □

The expressions for the covariance functions can be further characterized in some leading cases:

(1) The distributions of the covariates in groups 0 and 1 correspond to different populations and are estimated by the empirical distributions using mutually independent random

samples. In this case G_{X_0} and G_{X_1} are independent integrals over Brownian bridges, and the second component of the covariance function in (D.2) is $\int_{\mathcal{X}}[F_{Y_j}(y|x)-F^k_{Y_j}(y)][F_{Y_r}(\tilde{y}|x)-F^k_{Y_r}(\tilde{y})]dF_{X_k}(x)$ for $k=s$ and zero for $k\neq s$.

(2) The covariates in group j are known transformations of the covariates in group 0, $X_1=g(X_0)$, and the covariate distribution in group 0 is estimated by the empirical distribution from a random sample. In this case G_{X_0} and G_{X_1} are highly dependent processes. The second components of the covariance function in (D.2) is $\int_{\mathcal{X}}[F_{Y_j}(y|x)-F^0_{Y_j}(y)]\ [F_{Y_r}(\tilde{y}|x)-F^0_{Y_r}(\tilde{y})]dF_{X_0}(x)$ for $k=s=0$, $\int_{\mathcal{X}}[F_{Y_j}(y|g(x))-F^1_{Y_j}(y)]F_{Y_r}(\tilde{y}|g(x))-F^1_{Y_r}(\tilde{y})]dF_{X_0}(x)$ for $k=s=1$, and $\int_{\mathcal{X}}[F_{Y_j}(y|x)-F^0_{Y_j}(y)][F_{Y_r}(\tilde{y}|g(x))-F^1_{Y_r}(\tilde{y})]dF_{X_0}(x)$ for $k\neq s$.

Corollary 4. *Limit distribution theory and validity of bootstrap and other simulation methods for the estimators of the marginal quantile function, quantile policy effects, distribution policy effects, and differentiable functionals can be obtained using similar arguments to Theorems 2 and 3, and Corollaries 1–3 with obvious changes of notation.*

D.2. **Hadamard derivatives of marginal functionals.** In order to state the next result, we define the pseudometric $\rho_{L^2(P)}$ on $\mathcal{Y}\times\mathcal{X}$, and on $\mathcal{F}$ by

$$\rho^j_{L^2(P)}((y,x),(\tilde{y},\tilde{x}))=\left[E\Big\{Z_j(y,x)-Z_j(\tilde{y},\tilde{x})\Big\}^2\right]^{1/2},\ \text{for } j\in\{0,1\},$$

$$\rho^k_{L^2(P)}(f,\tilde{f})=\left[E\left\{G_{X_k}(f)-G_{X_k}(\tilde{f})\right\}^2\right]^{1/2},\ \text{for } k\in\{0,1\}.$$

It follows from Lemma 18.15 in van der Vaart (1998) that $\mathcal{Y}\times\mathcal{X}$ is totally bounded under $\rho^j_{L^2(P)}$ and Z_j has continuous paths with respect to $\rho^j_{L^2(P)}$ for each j. Moreover, the completion of $\mathcal{Y}\times\mathcal{X}$, denoted $\overline{\mathcal{Y}\times\mathcal{X}}$, with respect to either of the pseudometrics is compact. Likewise, $\mathcal{F}$ is totally bounded under $\rho^k_{L^2(P)}$ for each k.

Lemma 10. *Consider the mapping* $\phi:\mathbb{D}_\phi\subset\mathbb{D}=\ell^\infty(\mathcal{YX})\times\ell^\infty(\mathcal{F})\mapsto\mathbb{E}=\ell^\infty(\mathcal{Y})$,

$$\phi(F_{Y_j},F_{X_k}):=\int F_{Y_j}(\cdot|x)dF_{X_k}(x),\ j,k\in\{0,1\},$$

where the domain $\mathbb{D}_\phi$ is the product of the space of the conditional distribution functions $F_{Y_j}(\cdot|\cdot)\in\mathcal{F}$ on $\mathcal{YX}$ and the space of bounded maps $f\mapsto\int f dF_{X_k}$, where F_{X_k} is a distribution function on $\mathcal{X}$, for $j,k\in\{0,1\}$.[18] *Consider the sequence $(F^t_{Y_j},F^t_{X_k})\in\mathbb{D}_\phi$ such*

[18]That is, we identify F_{X_k} with the map $f\mapsto\int f dF_{X_k}$ in $\ell^\infty(\mathcal{F})$.

that for $\alpha_j^t := (F_{Y_j}^t - F_{Y_j})/(t\sqrt{\lambda_j})$, $d\beta_k^t := d(F_{X_k}^t - F_{X_k})/(t\sqrt{\lambda_k})$, *and* $\beta_k^t(f) := \int f d\beta_k^t$, *as* $t \searrow 0$

$$\alpha_j^t \to \alpha_j \in C(\mathcal{YX}, \rho_{L^2(P)}^j) \quad \text{in } \ell^\infty(\mathcal{YX}),$$
$$\beta_k^t \to \beta_k \in C(\mathcal{F}, \rho_{L^2(P)}^k) \quad \text{in } \ell^\infty(\mathcal{F}),$$

for the F_{X_k}*-Donsker class* $\mathcal{F}$ *and* $j, k \in \{0, 1\}$. *Finally, we assume that* $\{F_{Y_j}(y|x), y \in \mathcal{Y}\}$ *is* F_{X_k}*-Donsker, for* $j, k \in \{0, 1\}$. *Then, as* $t \searrow 0$

$$\frac{\phi(F_{Y_j}^t, F_{X_k}^t) - \phi(F_{Y_j}, F_{X_k})}{t} \to \phi'_{F_{Y_j}, F_{X_k}}(\alpha_j, \beta_k),$$

where

$$\phi'_{F_{Y_j}, F_{X_k}}(\alpha_j, \beta_k) := \sqrt{\lambda_j} \int \alpha_j(\cdot|x) dF_{X_k}(x) + \sqrt{\lambda_k} \int F_{Y_j}(\cdot|x) d\beta_k(x),$$

and the derivative map $(\alpha, \beta) \mapsto \phi'_{F_{Y_j}, F_{X_k}}(\alpha, \beta)$, *mapping* $\mathbb{D}_\phi$ *to* $\mathbb{E}$, *is continuous.*

Proof of Lemma 10. Write $\frac{\phi(F_{Y_j}^t, F_{X_k}^t) - \phi(F_{Y_j}, F_{X_k})}{t} - \phi'_{F_{Y_j}, F_{X_k}}(\alpha_j, \beta_k)$ as

$$\sqrt{\lambda_j} \int (\alpha_j^t - \alpha_j) dF_{X_k} + \sqrt{\lambda_k} \int F_{Y_j}(d\beta_k^t - d\beta_k) + \sqrt{\lambda_j \lambda_k} \int \alpha_j t d\beta_k^t + \sqrt{\lambda_j \lambda_k} \int (\alpha_j^t - \alpha_j) t d\beta_k^t \tag{D.3}$$

The first term of (D.3) is bounded by $\|\alpha_j^t - \alpha_j\|_{\mathcal{YX}} \int dF_{X_k} \to 0$. The second term vanishes, since for any F_{X_k}-Donsker set $\mathcal{F}$, $\int f d\beta_k^t \to \int f d\beta_k$ in $\ell^\infty(\mathcal{F})$, and $\{F_{Y_j}(y|x), y \in \mathcal{Y}\} \subset \mathcal{F}$ by assumption. The third term vanishes by the argument provided below. The fourth term vanishes, since $|\int (\alpha_j^t - \alpha_j) t d\beta_k^t| \le \|\alpha_j^t - \alpha_j\|_{\mathcal{YX}} \int |t d\beta_k^t| \le 2\|\alpha_j^t - \alpha_j\|_{\mathcal{YX}} \to 0$.

Since α_j is continuous on the compact semi-metric space $(\overline{\mathcal{YX}}, \rho_{L^2(P)}^j)$, there exists a finite measurable partition $\cup_{i=1}^m \mathcal{YX}_{im}$ of $\overline{\mathcal{YX}}$ such that α_j varies less than ϵ on each subset. 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 $1_{im}(y, x) = 1\{(y, x) \in \mathcal{YX}_{im}\}$. Then

$$\begin{aligned}
\left| \int \alpha_j t d\beta_k^t \right| &\le 2\|\alpha_j - \alpha_j \circ \pi_m\|_{\mathcal{YX}} + \sum_{i=1}^m |\alpha_j(y_{im}, x_{im})| t\beta_k^t(1_{im}) \\
&\le 2\epsilon + \sum_{i=1}^m |\alpha_j(y_{im}, x_{im})| t(\beta_k(1_{im} + o(1)) \\
&\le 2\epsilon + tm \left[\|\alpha_j\|_{\mathcal{YX}} \max_{i \le m} \beta_k(1_{im}) + o(1) \right], \\
&\le 2\epsilon + O(t),
\end{aligned}$$

since $\{1_{im}, i \le m\}$ is a F_{X_k}-Donsker class. The constant ϵ is arbitrary, so the left hand side of the preceding display converges to zero.

Finally, the norm on $\mathbb{D}$ is 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 $\|\cdot\|_{\mathcal{YX}}$ by the first term in (D.3) vanishing, as shown above. Hence the derivative map is continuous. □

Appendix E. Functional Delta Method and Bootstrap and Other Simulation Methods for Z-processes

This section derives a preliminary result that is key to deriving the limit distribution and inference theory for various estimators of the conditional distribution and quantile functions. This result shows that suitably defined Z-estimators satisfy a functional central limit theorem and that we can estimate their laws using bootstrap and related methods. The result follows from a lemma that establishes Hadamard differentiability of Z-functionals in spaces that are particularly well-suited for our applications.

E.1. **Limit distribution and inference theory for approximate Z-processes.** Let us consider an index set T and a set $\Theta \subset \mathbb{R}^p$. We consider Z-estimation processes $\{\widehat{\theta}(u), u \in T\}$, where for each $u \in T$, $\widehat{\theta}(u)$ satisfies $\|\widehat{\Psi}(\widehat{\theta}(u), u)\| \leq \inf_{\theta\in\Theta} \|\widehat{\Psi}(\theta, u)\| + \epsilon_n$, with $\epsilon_n \searrow 0$ at some rate. That is, $\widehat{\theta}(u)$ is an approximate solution to the problem of minimizing $\|\widehat{\Psi}(\theta, u)\|$ over $\theta \in \Theta$. The random function $(\theta, u) \mapsto \widehat{\Psi}(\theta, u)$ is an estimator of some fixed population function $(\theta, u) \mapsto \Psi(\theta, u)$, and satisfies a functional central limit theorem. The following lemma specifies conditions under which the Z-processes satisfy a functional central limit theorem, and under which bootstrap and other simulation methods consistently estimate the law of this process.

Lemma 11 (Limit distribution and inference theory for approximate Z-processes)**.** *Let T be a relatively compact set of some metric space, and Θ be a compact subset of $\mathbb{R}^p$. Assume that*

(i) *for each $u \in T$, $\Psi(\cdot, u) : \Theta \mapsto \mathbb{R}^p$ possesses a unique zero at $\theta_0(u) \in$ interior Θ, and has inverse $\Psi^{-1}(\cdot, u)$ that is continuous at 0 uniformly in $u \in T$,*

(ii) *$\Psi(\cdot, u)$ is continuously differentiable at $\theta_0(u)$ uniformly in $u \in T$, with derivative $\dot{\Psi}_{\theta_0(u),u}$ that is uniformly non-singular, namely $\inf_{u\in T} \inf_{\|h\|=1} \|\dot{\Psi}_{\theta_0(u),u} h\| > 0$.*

(iii) *$\sqrt{n}(\widehat{\Psi} - \Psi) \Rightarrow Z$ in $\ell^\infty(\Theta \times T)$, where Z is a.s. continuous on $\Theta \times T$ with respect to the Euclidean metric,*

(iv) *Bootstrap or some other method consistently estimates the law of $\sqrt{n}(\widehat{\Psi} - \Psi)$.*

For each $u \in T$, let $\widehat{\theta}(u)$ be such that $\|\widehat{\Psi}(\widehat{\theta}(u), u)\| \leq \inf_{\theta\in\Theta} \|\widehat{\Psi}(\theta, u)\| + \epsilon_n$, with $\epsilon_n = o(n^{-1/2})$. Then, under conditions (i)–(iii)

$$\sqrt{n}(\widehat{\theta}(\cdot) - \theta_0(\cdot)) \Rightarrow -\dot{\Psi}^{-1}_{\theta_0(\cdot),\cdot}[Z(\theta_0(\cdot), \cdot)] \quad \text{in } \ell^\infty(T).$$

Moreover, any bootstrap or other method that satisfies condition (iv) consistently estimates the law of the empirical process $\sqrt{n}(\widehat{\theta} - \theta_0)$ in $\ell^\infty(T)$.

Proof of Lemma 11. The results follow by the functional delta method in Lemma 3 and by the functional delta method for bootstrap and other methods in Lemma 6, and the Hadamard differentiability of Z-functionals established in Lemma 12 with $t = 1/\sqrt{n}$. □

The proof of the preceding result relies on the following lemma. Let T be a relatively compact set of some metric space, and Θ be a compact subset of $\mathbb{R}^p$. An element $\theta \in \Theta$ is an r-approximate zero of the map $\theta \mapsto z(\theta, u)$ if for some $r > 0$

$$\|z(\theta, u)\| \leq \inf_{\theta'\in\Theta} \|z(\theta', u)\| + r.$$

Let $\phi(\cdot, r) : \ell^\infty(\Theta) \mapsto \Theta$ be a map that assigns one of its r-approximate zeroes $\phi(z(\cdot, u), r)$ to each element $z(\cdot, u) \in \ell^\infty(\Theta)$.

Lemma 12. *Assume that conditions (i) and (ii) on the function Ψ stated in the preceding lemma hold. Take any $z_t \to z$ uniformly on $\Theta \times T$ as $t \searrow 0$, for a continuous map $z : \Theta \times T \mapsto \mathbb{R}^p$, and suppose that $q_t \searrow 0$ uniformly on T as $t \searrow 0$. Then, for the $tq_t(u)$-approximate zero of $\Psi(\cdot, u) + tz_t(\cdot, u)$ denoted as $\theta_t(u) = \phi(\Psi(\cdot, u) + tz_t(\cdot, u), tq_t(u))$ we have that, uniformly in $u \in T$,*

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

Here it is useful to think of t as $1/\sqrt{n}$, where n is the sample size.

Remark. Our lemma is an alternative to van der Vaart and Wellner's (1996) Lemma 3.9.34 on Hadamard differentiability of Z-functionals in general normed spaces. The conditions of their lemma are difficult to meet in our context because they include the uniform convergence of the functions z_t over the parameter space $\mathcal{F} = \ell^\infty(T)$, the collection of all bounded functions on T, which is an extremely large parameter space. In particular, to apply their lemma we need that the empirical processes $\sqrt{n}(\widehat{\Psi} - \Psi)$ indexed by $\mathcal{F} = \ell^\infty(T)$ converge weakly in the space $\ell^\infty(\mathcal{F} \times T)$, which appears to be difficult to attain in applications such as quantile regression processes. Indeed, note that weak convergence in this space requires $\mathcal{F}$ to be totally bounded, which is hard to attain when $\mathcal{F}$ is too rich a space.

See Van der Vaart and Wellner (1996) p. 396 for a comment on the limitation of their Lemma 3.9.34. Moreover, our lemma allows for approximate Z-estimators. This allows us to cover quantile regression processes, where exact Z-estimators do not exist.

Proof of Lemma 12. We have that $\Psi(\theta_0(u), u) = 0$ for all $u \in T$. Let $z_t \to z$ uniformly on $\Theta \times T$ for a map $z : \Theta \times T \mapsto \ell^\infty(\Theta \times T)$ that is continuous at each point, and $q_t \searrow 0$ uniformly in $u \in T$ as $t \searrow 0$. By definition $\theta_t(u) = \phi(\Psi(\cdot, u) + tz_t(\cdot, u), tq_t(u))$ satisfies

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

uniformly in $u \in T$. The the rest of the proof has three steps. In Step 1, we establish a rate of convergence for $\theta_t(\cdot)$ to $\theta(\cdot)$. In Step 2, we verify the main claim of the lemma concerning the linear representation for $t^{-1}(\theta_t(\cdot) - \theta(\cdot))$, 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 $u \in T$, $\|\theta_t(u) - \theta_0(u)\| \leq c^{-1}\|\Psi(\theta_t(u), u) - \Psi(\theta_0(u), u)\| = O(t)$. 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 $u \in T$. We conclude that uniformly in $u \in T$, 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 $u \in T$, $\|\Psi(\theta_t(u), u) - \Psi(\theta_0(u), u)\| = O(t)$. By assumption $\Psi(\cdot, u)$ has a unique zero at $\theta_0(u)$ and has an inverse that is continuous at zero uniformly in $u \in T$; hence it follows that uniformly in $u \in T$,

$$\|\theta_t(u) - \theta_0(u)\| \leq d_H(\Psi^{-1}(\Psi(\theta_t(u), u), u), \Psi^{-1}(0, u)) \to 0,$$

where d_H is the Hausdorff distance. By continuous differentiability assumed to hold uniformly in $u \in T$, $\|\Psi(\theta_t(u), u) - \Psi(\theta_0(u), u) - \dot{\Psi}_{\theta_0(u),u}[\theta_t(u) - \theta_0(u)]\| = o(\|\theta_t(u) - \theta_0(u)\|)$ so that uniformly in $u \in T$

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

where h ranges over $\mathbb{R}^p$, and $c > 0$ by assumption. Thus, uniformly in $u \in T$, $\|\theta_t(u) - \theta_0(u)\| \leq c^{-1}\|\Psi(\theta_t(u), u) - \Psi(\theta_0(u), u)\| = O(t)$.

Step 2. Here we verify the main claim of the lemma. Using continuous differentiability uniformly in u again, conclude $\|\Psi(\theta_t(u), u) - \Psi(\theta_0(u), u) - \dot{\Psi}_{\theta_0(u),u}[\theta_t(u) - \theta_0(u)]\| = o(t)$. Below we will show that $\lambda_t(u) = o(1)$ and we also have $q_t(u) = o(1)$ uniformly in $u \in$

T by assumption. Thus, we can conclude that uniformly in $u \in T$, $t^{-1}(\Psi(\theta_t(u), u) - \Psi(\theta_0(u), u)) = -z_t(\theta_t(u), u) + o(1) = -z(\theta_0(u), u) + o(1)$ 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 $u \in T$. 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 $\bar{\theta}_t \in \Theta$, for small enough t, uniformly in $u \in T$; moreover, $\lambda_t(u) \leq \|t^{-1}\Psi(\bar{\theta}_t(u), u) + z_t(\bar{\theta}_t(u), u)\| = \| - \dot{\Psi}_{\theta_0(u),u}\{\dot{\Psi}^{-1}_{\theta_0(u),u}[z(\theta_0(u), u)]\} + z(\theta_0(u), u) + o(1)\| = o(1)$, as $t \searrow 0$. □

APPENDIX F. Z-ESTIMATORS OF CONDITIONAL QUANTILE AND DISTRIBUTION FUNCTIONS

This section derives limit theory for the principal estimators of conditional distribution and quantile functions. These results establish the validity of bootstrap and other resampling plans for the entire quantile regression process, the entire distribution regression process, and related processes arising in estimation of various conditional quantile and distribution functions. These results may be of a substantial independent interest.

In order to prove the results, we use Lemmas 11 and 12. We also specify some primitive conditions that cover all of our leading examples. In all these examples, we have functional parameter values $u \mapsto \theta(u)$ where $u \in T \subset \mathbb{R}$ and $\theta(u) \subset \Theta \subset \mathbb{R}^p$, where for each $u \in T$, $\theta_0(u)$ solves the equation

$$\Psi(\theta, u) := E[g(W, \theta, u)] = 0,$$

where $g : \mathcal{W} \times \Theta \times T \mapsto \mathbb{R}^p$, $W := (X, Y)$ is a random vector with support $\mathcal{W}$. For estimation purposes we have an empirical analog of the above moment functions

$$\widehat{\Psi}(\theta, u) = E_n[g(W_i, \theta, u)]$$

where E_n is the empirical expectation and $(W_1, ..., W_n)$ is a random sample from W. For each $u \in T$, the estimator $\widehat{\theta}(u)$ satisfies $\|\widehat{\Psi}(\widehat{\theta}(u), u)\| \leq \inf_{\theta \in \Theta} \|\widehat{\Psi}(\theta, u)\| + \epsilon_n$, with $\epsilon_n = o(n^{-1/2})$.

CONDITION Z.1. *The set Θ is a compact subset of $\mathbb{R}^p$ and T is either a finite subset or a bounded open subset of $\mathbb{R}^d$.*

(i) *For each $u \in T$, $\Psi(\theta, u) := Eg(W, \theta, u) = 0$ has a unique zero at $\theta_0(u) := (\alpha_0(u)', \beta_0')' \in$ interior Θ.*

(ii) *The map* $(\theta, u) \mapsto \Psi(\theta, u)$ *is continuously differentiable at* $(\theta_0(u), u)$ *with a uniformly bounded derivative on* T*, where differentiability in* u *needs to hold for the case of* T *being a bounded open subset of* $\mathbb{R}^d$ *;* $\dot{\Psi}_{\theta,u} = G(\theta, u) = \frac{\partial}{\partial \theta'} Eg(W, \theta, u)$ *is uniformly nonsingular at* $\theta_0(u)$*, namely* $\inf_{u \in T} \inf_{\|h\|=1} \|\dot{\Psi}_{\theta_0(u),u} h\| > 0$.

(iii) *The function set* $\mathcal{G} = \{g(W, \theta, u), (\theta, u) \in \Theta \times T\}$ *is P-Donsker with a square integrable envelope* $\bar{G}$*. The map* $(\theta, u) \mapsto g(W, \theta, u)$ *is continuous at each* $(\theta, u) \in \Theta \times T$ *with probability one.*

Condition Z.2. *Either of the following holds:*

(a) *the conditional distribution has the form* $F_Y(u|x) = \Lambda(x, \theta_0(u))$*; or*

(b) *the quantile functions have the form* $Q_Y(u|x) = Q(x, \theta_0(u))$*, where the functions* $\theta \mapsto \Lambda(x, \theta)$ *and* $\theta \mapsto Q(x, \theta)$ *are continuously differentiable in* θ *with derivatives that are uniformly bounded over the set* $\mathcal{X}$*.*

Lemma 13. *Condition Z.1 implies conditions (i)-(iv) of Lemma 11. In particular, condition (iii) holds with* $\sqrt{n}(\widehat{\Psi} - \Psi) \Rightarrow Z$*, in* $\ell^\infty(T)$*, where* Z *is a zero mean Gaussian process with continuous paths in* $u \in T$ *and covariance function*

$$\Omega(u, \tilde{u}) = E[g(W, \theta_0(u), u) g(W, \theta_0(\tilde{u}), \tilde{u})'].$$

Condition (iv) holds with the set of consistent methods for estimating the law of $\sqrt{n}(\widehat{\Psi} - \Psi)$ *consisting of bootstrap and exchangeable bootstraps, more generally. Consequently, the conclusions of Lemma 11 hold, namely* $\sqrt{n}(\widehat{\theta}(\cdot) - \theta_0(\cdot)) \Rightarrow -G(\theta_0(\cdot), \cdot)^{-1} [Z(\theta_0(\cdot), \cdot)]$ *in* $\ell^\infty(T)$*. Moreover, bootstrap and exchangeable bootstraps consistently estimate the law of the empirical process* $\sqrt{n}(\widehat{\theta} - \theta_0)$*.*

This lemma presents a useful result in its own right. From the point of view of this paper, the following result, a corollary of the lemma, is of immediate interest to us since it verifies Condition D and Condition Q for a wide class of estimators of conditional distribution and quantile functions.

Theorem 6 (Limit distribution and inference theory for Z-estimators of conditional distribution and quantile functions)**.** *1. Under conditions Z.1-Z.2(a), the estimator* $(u, x) \mapsto \widehat{F}_Y(u|x)$ *of the conditional distribution function* $(u, x) \mapsto F_Y(u|x)$ *converges in law to a continuous Gaussian process:*

$$\sqrt{n} \left(\widehat{F}_Y(u|x) - F_Y(u|x) \right) \Rightarrow Z(u, x) := -\frac{\partial \Lambda(x, \theta_0(u))}{\partial \theta'} G(\theta_0(u), u)^{-1} Z(\theta_0(u), u) \quad \text{(F.1)}$$

in $\ell^\infty(\mathcal{Y}\times\mathcal{X})$*, where* $(u,x)\mapsto Z(u,x)$ *has zero mean and covariance function* $\Sigma_Z(u,x,\tilde{u},\tilde{x}) := E[Z(u,x)Z(\tilde{u},\tilde{x})]$*. Moreover, bootstrap and exchangeable bootstraps consistently estimate the law of* Z*.*

2. Under conditions Z.1-Z.2(b), the estimator $(u,x)\mapsto \widehat{Q}_Y(u|x)$ *of the conditional quantile function* $(u,x)\mapsto Q_Y(u|x)$ *converges in law to a continuous Gaussian process:*

$$\sqrt{n}\left(\widehat{Q}_Y(u|x) - Q_Y(u|x)\right) \Rightarrow V(u,x) := -\frac{\partial Q(x,\theta(u))}{\partial\theta'} G(\theta_0(u),u)^{-1} Z(\theta_0(u),u), \quad \text{(F.2)}$$

in $\ell^\infty((0,1)\times\mathcal{X})$*, where the process* $(u,x)\mapsto V(u,x)$ *has zero mean and covariance function* $\Sigma_V(u,x,\tilde{u},\tilde{x}) := E[V(u,x)V(\tilde{u},\tilde{x})]$*. Moreover, bootstrap and exchangeable bootstraps consistently estimate the law of* V*.*

Proof of Lemma 13. We shall verify conditions (i)-(iv) of Lemma 11.

We consider the case where T is a bounded open subset of $\mathbb{R}$. The proof for the case with a finite T is simpler, and follows similarly. To show condition (i), we note that by the implicit function theorem and uniqueness of θ_0, the inverse map $\Psi^{-1}(\mu,u)$ exists on a open neighborhood of each pair $(\mu=0,u)$, and it is continuously differentiable in (μ,u) at each pair $(\mu=0,u)$ with a uniformly bounded derivative. This implies that for any sequence of points $(\mu_t,u_t)\to(0,u)$ with $u\in\bar{T}$, where $\bar{T}$ is the closure of T, we have that $\|\Psi^{-1}(\mu_t,u_t) - \Psi^{-1}(0,u_t)\| = O(\|\mu_t\|) = o(1)$, verifying the continuity of the inverse map at 0 uniformly in u. We can also conclude that $\theta_0(u) = \Psi^{-1}(0,u)$ is uniformly continuous on T and we can extend it to $\bar{T}$ by taking limits.

To show condition (ii) we take any sequence $(u_t,h_t)\to(u,h)$ with $u\in T, h\in\mathbb{R}^p$ and then note that, for $t^*\in[0,t]$

$$\begin{aligned}\Delta_t(u_t,h_t) &= t^{-1}\{\Psi(\theta_0(u_t)+th_t,u_t) - \Psi(\theta_0(u_t),u_t)\} = \frac{\partial\Psi}{\partial\theta}(\theta_0(u_t)+t^*h_t,u_t)h_t \\ &\to \frac{\partial\Psi}{\partial\theta}(\theta_0(u),u)h = G(\theta_0(u),u)h,\end{aligned}$$

using the continuity hypotheses on the derivative $\partial\Psi/\partial\theta$ and the continuity of $u\mapsto\theta_0(u)$. Hence by Lemma 5, we conclude that $\sup_{u\in T,\|h\|=1}|\Delta_t(u,h) - G(\theta_0(u),u)h|\to 0$ as $t\searrow 0$.

To show condition (iii), note that by the Donsker central limit theorem for $\widehat{\Psi}(\theta,u) = \mathbb{E}_n[g(W_i,\theta,u)]$ we have that $\sqrt{n}(\widehat{\Psi}-\Psi)\Rightarrow Z$, where Z is a zero mean Gaussian process with covariance function $\Omega(u,\tilde{u}) = E[g(W,\theta_0(u),u)g(W,\theta_0(\tilde{u}),\tilde{u})']$ that has continuous paths with respect to the $L_2(P)$ semi-metric on $\mathcal{G}$. The map $(\theta,u)\mapsto g(W,\theta,u)$ is continuous at each (θ,u) with probability one. The only result that is not immediate from the assumptions stated is that Z also has continuous paths on $\Theta\times T$ with respect

to the Euclidean metric $\|\cdot\|$. By assumption Z has continuous paths with respect to $\rho_{L^2(P)}((\theta,u),(\tilde{\theta},\tilde{u})) = \{E[g(W,\theta,u) - g(W,\tilde{\theta},\tilde{u})]^2\}^{1/2}$. As $\|(\theta,u)-(\tilde{\theta},\tilde{u})\| \to 0$, we have that $g(W,\theta,u) - g(W,\tilde{\theta},\tilde{u}) \to 0$ almost surely. It follows by the dominated convergence theorem, with dominating function equal to $(2\bar{G})^2$, where $\bar{G}$ is the square integrable envelope for the function class $\mathcal{G}$, that $\{E[g(W,\theta,u) - g(W,\tilde{\theta},\tilde{u})]^2\}^{1/2} \to 0$. This verifies the continuity condition. The square integrable envelope $\bar{G}$ exists by assumption.

To show (iv), we simply invoke Theorem 3.6.13 in Van der Vaart and Wellner (1996) which implies that the bootstrap and exchangeable bootstraps, more generally, consistently estimate the limit law of $\sqrt{n}(\widehat{\Psi} - \Psi)$, say $\mathbb{G}$, in the sense of equation (A.2). □

Proof of Theorem 6. This result follows directly from Lemma 12, the functional delta method in Lemma 3, the chain rule for Hadamard differentiable functionals in Lemma 4, and the preservation of validity of bootstrap and other methods for Hadamard differentiable functionals in Lemma 6. □

F.1. **Examples of conditional quantile estimation methods.** We consider the location and quantile regression models described in the text.

Example 2. Quantile regression. The conditional quantile function of the outcome variable Y given the covariate vector X is given by $X'\beta_0(\cdot)$. Here we can take the moment functions corresponding to the canonical quantile regression approach:

$$g(W,\beta,u) = (u - 1\{Y \leq X'\beta\})X. \tag{F.3}$$

We assume that the conditional density $f_Y(\cdot|X)$ is uniformly bounded and is continuous at $X'\beta_0(u)$ uniformly in $u \in T$, almost surely; moreover, $\inf_{u\in T} f_Y(X'\beta_0(u)|X) \geq c > 0$ almost surely; and $E[XX']$ is finite and of full rank. The true parameter $\beta_0(u)$ solves $Eg(W,\beta,u) = 0$ and we assume that the parameter space Θ is such that $\beta_0(u) \in$ interior Θ for each $u \in (0,1)$.

Lemma 14. *Conditions Z.1-Z.2(b) hold for this example with moment function given by (F.3), $T = (0,1)$, $Q_Y(u|x) = x'\beta_0(u)$, $G(\beta_0(u),u) = -E[f_Y(X'\beta_0(u)|X)XX']$, and $\Omega(u,\tilde{u}) = \{\min(u,\tilde{u}) - u\tilde{u}\}E[XX']$.*

Proof of Lemma 14. To show Z.1, we need to verify conditions on the derivatives of the map $\beta \mapsto Eg(W,\beta,u)$. It is straightforward to show that we have that at $(\beta,u) = (\beta_0(u),u)$,

$$\frac{\partial}{\partial(\beta',u)} Eg(W,\beta,u) = [G(\beta,u), EX] = [-E[f_Y(X'\beta|X)XX'], EX],$$

and the right hand side is continuous at $(\beta_0(u), u)$. This follows using the dominated convergence theorem, the a.s. continuity and boundedness of the mapping $y \mapsto f_Y(y|X)$ at $X'\beta_0(u)$, as well as finiteness of $E\|X\|^2$. Finally, note that $\beta_0(u)$ is the unique solution to $Eg(W, \beta, u) = 0$ for each u because it is a root of a gradient of convex function. Moreover, uniformly in $u \in (0, 1)$, $G(\beta_0(u), u) \geq \underline{f} EXX' > 0$, where $\underline{f}$ is the uniform lower bound on $f_Y(X'\beta_0(u)|X)$.

To show Z.1(iii) we verify that the function class $\mathcal{G}$ is P-Donsker with a square integrable envelope and the continuity hypothesis. The function classes $\mathcal{F}_1 = T$, $\mathcal{F}_2 = 1\{Y \leq X'\beta, \beta \in \mathbb{R}^p\}$ are VC classes. Therefore the function classes $\mathcal{F}_{kj} = \mathcal{F}_k X_j$ are also VC classes because they are formed as products of a VC class with a fixed function (Lemma 2.6.18 in van der Vaart and Wellner, 1996). The difference $\mathcal{F}_{1j} - \mathcal{F}_{2j}$ is a Lipschitz transform of VC classes, so it is P-Donkser by Example 19.9 in van der Vaart, 1998. The collection $\mathcal{G} = \{\mathcal{F}_{1j} - \mathcal{F}_{2j}, j = 1, ..., p\}$ is thus also Donsker. The envelope is given by $2 \max_j |X_j|$ which is square-integrable. Finally, the map $(\theta, u) \mapsto (u - 1(Y \leq X'\beta))X$ is continuous at each $(\beta, u) \in \Theta \times T$ with probability one by the absolute continuity of the conditional distribution of Y.

To show Z.2(b), we note that the map $(x, \theta) \mapsto x'\theta$ trivially verifies the hypotheses of Z.2(b) provided the set $\mathcal{X}$ is compact. □

Example 1. Classical regression. This is the location model $Y = X'\beta_0 + V$, where X is independent of V, so the conditional quantile function of outcome variable Y given the conditioning variable X is given by $X'\beta_0 + \alpha_0(\cdot)$, where $E[Y|X] = X'\beta_0$ and $\alpha_0(\cdot) = Q_V(\cdot)$. Here we can take the moment functions corresponding to using least squares to estimate β_0 and sample quantiles of residuals to estimate α_0.

$$g(W, \alpha, \beta, u) = [(u - 1\{Y - X'\beta \leq \alpha\}), (Y - X'\beta)X']'. \tag{F.4}$$

We assume that the density of $V = Y - X'\beta_0$, $f_V(\cdot)$ is uniformly bounded and is continuous at $\alpha_0(u)$ uniformly in $u \in T$, almost surely; moreover, $\inf_{u \in T} f(\alpha_0(u)) \geq c > 0$ almost surely; EXX' is finite, and full rank, and $EY^2 < \infty$. The true parameter value $(\alpha_0(u), \beta_0')'$ solves $Eg(W, \alpha, \beta, u) = 0$ and we assume that the parameter space Θ is such that $(\alpha_0(u), \beta_0')' \in$ interior Θ for each $u \in (0, 1)$.

Lemma 15. *Conditions Z.1-Z.3(b) hold for this example with moment function given by (F.4), $T = (0, 1)$, $Q_Y(u|x) = x'\beta_0 + \alpha_0(u)$,*

$$G(\alpha_0(u), \beta_0, u) = -\begin{bmatrix} f_V(\alpha_0(u)) & f_V(\alpha_0(u))E[X]' \\ 0_{p\times 1} & EXX' \end{bmatrix}, \tag{F.5}$$

and

$$\Omega(u,\tilde{u}) = \begin{bmatrix} \min(u,\tilde{u}) - u\tilde{u} & -E[V\ 1\{V \leq \alpha_0(u)\}]E[X]' \\ -E[V\ 1\{V \leq \alpha_0(u)\}]E[X] & E[V^2]EXX' \end{bmatrix}. \tag{F.6}$$

Proof of Lemma 15. The proof follows analogously to the proof of Lemma 14. Uniqueness of roots can also be argued similarly, with β_0 uniquely solving the least squares normal equation, and α_0 uniquely solving the quantile equation. □

F.2. **Examples of conditional distribution function estimation methods.** We consider the distribution regression model described in the text and an alternative estimator for the duration model based on distribution regression.

Example 4. Distribution regression. The conditional distribution function of the outcome variable Y given the covariate vector X is given by $\Lambda(X'\beta_0(\cdot))$, where Λ is either the probit or the logit link function. Here we can take the moment functions corresponding to the pointwise maximum likelihood estimation:

$$g(W,\beta,y) = \frac{\Lambda(X'\beta) - 1\{Y \leq y\}}{\Lambda(X'\beta)(1-\Lambda(X'\beta))}\lambda(X'\beta)X, \tag{F.7}$$

where λ is the derivative of Λ. Let $\mathcal{Y}$ be either a finite set or a bounded open subset of $\mathbb{R}^d$. For the latter case we assume that the conditional distribution function $y \mapsto F_Y(y|X)$ admits a density $y \mapsto f_Y(y|x)$, which is continuous at each $y \in \mathcal{Y}$, a.s. Moreover, EXX' is finite and full rank; the true parameter value $\beta_0(y)$ belongs to the interior of the parameter space Θ for each $y \in \mathcal{Y}$; and $\Lambda(X'\beta)(1-\Lambda(X'\beta)) \geq c > 0$ uniformly on $\beta \in \Theta$, a.s.

Lemma 16. *Conditions Z.1-Z.2(a) hold for this example with moment function given by (F.7), $T = \mathcal{Y}$, $u = y$, $F_Y(y|x) = \Lambda(x'\beta_0(y))$,*

$$G(\beta_0(y),y) := E\left[\frac{\lambda(X'\beta_0(y))^2}{\Lambda(X'\beta_0(y))[1-\Lambda(X'\beta_0(y))]}XX'\right],$$

and, for $\tilde{y} \geq y$,

$$\Omega(y,\tilde{y}) = E\left[\frac{\lambda(X'\beta_0(y))\lambda(X'\beta_0(\tilde{y}))}{\Lambda(X'\beta_0(y))[1-\Lambda(X'\beta_0(\tilde{y}))]}XX'\right].$$

Proof of Lemma 16. We consider the case where $\mathcal{Y}$ is a bounded open subset of $\mathbb{R}^d$. The case where $\mathcal{Y}$ is a finite set is simpler and follows similarly.

To show Z.1, we need to verify conditions on the derivatives of the map $\beta \mapsto Eg(W,\beta,u)$. By a straightforward calculation we have that at $(\beta,y) = (\beta_0(y),y)$,

$$\frac{\partial}{\partial(\beta',y)}Eg(W,\beta,y) = \left[E[\frac{\partial}{\partial\beta'}g(W,\beta,y)], [\frac{\partial}{\partial y}Eg(W,\beta,y)]\right] = [G(\beta,y), R(\beta,y)],$$

where, for $H(z) = \lambda(z)/\{\Lambda(z)[1-\Lambda(z)]\}$ and $h(z) = dH(z)/dz$,

$$\begin{aligned} G(\beta, y) &:= E\left[\{h(X'\beta)[\Lambda(X'\beta) - 1\{Y \leq y\}] + H(X'\beta)\lambda(X'\beta)\}XX'\right], \\ R(\beta, y) &= E\left[H(X'\beta)f_Y(y|X)X\}\right]. \end{aligned}$$

Both terms are continuous in (β, y) at $(\beta_0(y), y)$ for each $y \in \mathcal{Y}$. This follows from using by the dominated convergence theorem and the following ingredients: (1) a.s. continuity of the map $(\beta, y) \mapsto \frac{\partial}{\partial \beta'} g(W, \beta_0(y), y)$, (2) domination of $\|\frac{\partial}{\partial \beta'} g(W, \beta, y)\|$ by a square-integrable function const$\|X\|$, (3) a.s. continuity of the conditional density function $y \mapsto f_Y(y|X)$, and (4) $\Lambda(X'\beta)(1-\Lambda(X'\beta)) \geq c > 0$ uniformly on $\beta \in \Theta$, a.s. Finally, also note that the solution $\beta_0(y)$ to $Eg(W, \beta, y) = 0$ is unique for each $y \in \mathcal{Y}$ because it is a root of a gradient of a convex function.

To show Z.1(iii), we verify that the function class $\mathcal{G}$ is P-Donsker with a square integrable envelope. Function classes $\mathcal{F}_1 = \{X'\beta, \beta \in \Theta\}$, $\mathcal{F}_2 = \{1\{Y \leq y\}, y \in \mathcal{Y}\}$, and $\{X_j\}$, $j = 1, ..., p$ are VC classes of functions. The final class

$$\mathcal{G} = \left\{ \frac{\Lambda(\mathcal{F}_1) - \mathcal{F}_2}{\Lambda(\mathcal{F}_1)(1-\Lambda(\mathcal{F}_1))} \lambda(\mathcal{F}_1) X_j, \quad j = 1, ..., p \right\},$$

is a Lipschitz transformation of VC classes with Lipschitz coefficient bounded by $c \max_j |X_j|$ and the envelope function $c' \max_j |X_j|$, which are square-integrable; here 1 and c' are some positive constants. Hence $\mathcal{G}$ is Donsker by Example 19.9 in van der Vaart (1998). Finally, the map

$$(\beta, y) \mapsto \frac{\Lambda(X'\beta) - 1\{Y \leq y\}}{\Lambda(X'\beta)(1-\Lambda(X'\beta))} \lambda(X'\beta) X$$

is continuous at each $(\beta, y) \in \Theta \times \mathcal{Y}$ with probability one by the absolute continuity of the conditional distribution of Y and by the assumption that $\Lambda(X'\beta)(1-\Lambda(X'\beta)) \geq c > 0$ uniformly on $\beta \in \Theta$, a.s.

To show Z.2(a), we note that the map $(x, \theta) \mapsto \Lambda(x'\theta)$ trivially verifies the hypotheses of Z.2(a) provided the set $\mathcal{X}$ is compact. □

Example 3b. Duration regression. An alternative to the proportional hazard model in duration and survival analysis is to specify the conditional distribution function of the duration Y given the covariate vector X as $\Lambda(\alpha_0(\cdot) + X'\beta_0)$, where Λ is either the probit or the logit link function. We normalize $\alpha_0(y_0) = 0$ at some $y_0 \in \mathcal{Y}$. Here we can

take the following moment functions:

$$g(W,\alpha,\beta,y) = \begin{bmatrix} \dfrac{\Lambda(\alpha+X'\beta)-1\{Y\le y\}}{\Lambda(\alpha+X'\beta)(1-\Lambda(\alpha+X'\beta))}\lambda(\alpha+X'\beta) \\ \dfrac{\Lambda(X'\beta)-1\{Y\le y_0\}}{\Lambda(X'\beta)(1-\Lambda(X'\beta))}\lambda(X'\beta)X \end{bmatrix}$$

where λ is the derivative of Λ. The first set of equations is used for estimation of $\alpha_0(y)$ and the second for estimation of β_0.

Let $\mathcal{Y}$ be either a finite set or a bounded open subset of $\mathbb{R}^d$. For the latter case we assume that the conditional distribution function $y \mapsto F_Y(y|X)$ admits a density $y \mapsto f_Y(y|x)$, which is continuous at each $y \in \mathcal{Y}$, a.s. Moreover, EXX' is finite and full rank; the true parameter value $(\alpha_0(y), \beta_0')'$ belongs to the interior of the parameter space Θ for each $y \in \mathcal{Y}$; and $\Lambda(\alpha + X'\beta)(1 - \Lambda(\alpha + X'\beta)) \ge c > 0$ uniformly on $(\alpha, \beta')' \in \Theta$, a.s.

Lemma 17. *Conditions Z.1-Z.2(a) hold for this example with moment function given by (F.7), $T = \mathcal{Y}$, $u = y$, $F_Y(y|x) = \Lambda(\alpha_0(y) + x'\beta_0)$,*

$$G(\alpha_0(y), \beta_0, y) = E\frac{\partial}{\partial(\alpha,\beta')}g(W, \alpha_0(y), \beta_0),$$

and $\Omega(y, \tilde{y}) = E[g(W, \alpha_0(y), \beta_0)g(W, \alpha_0(\tilde{y}), \beta_0)']$.

Proof of Lemma 17. The proof follows analogously to the proof of Lemma 16. □

References

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

[2] Abadie, A., Angrist, J., and G. Imbens (2002): "Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings," *Econometrica* 70, pp. 91–117.

[3] Andersen, P. K., and R. D. Gill (1982): "Cox's Regression Model for Counting Processes: A Large Sample Study," *The Annals of Statistics* 10, pp. 1100–1120.

[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 (2003): "Consistent Tests for Stochastic Dominance," *Econometrica*, 71, pp. 71–104.

[10] 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.

[11] Beran, R.J. (1977): "Estimating a distribution function," *Annals of Statistics* 5, pp. 400-404.

[12] Breslow, N. E. (1972): Contribution to the Discussion of "Regression Models and Life Tables," by D. R. Cox, *Journal of the Royal Statistical Society*, Ser. B, 34, pp. 216–217.

[13] Breslow, N. E. (1974): "Covariance Analysis of Censored Survival Data," *Biometrics*, 30, pp. 89–100.

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

[15] 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.

[16] Cameron, A.C., and P.K. Trivedi (2005): *Microeconometrics: Methods and Applications,* Cambridge University Press, New York.

[17] 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.

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

[19] Chernozhukov, V., I. Fernandez-Val and A. Galichon (2006): "Quantile and Probability Curves without Crossing," mimeo, 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] Chesher, A. (2003): "Identification in Nonseparable Models." *Econometrica,* 71, pp. 1405–1441.

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

[25] Dabrowska, D.M. (2005): "Quantile Regression in Transformation Models," *Sankhyā* 67, pp. 153–186.

[26] DiNardo, J. (2002): "Propensity Score Reweighting and Changes in Wage Distributions," *unpublished manuscript*, University of Michigan.

[27] 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.

[28] Doksum, K. (1974): "Empirical Probability Plots and Statistical Inference for Nonlinear Models in the Two-Sample Case," *Annals of Statistics* 2, pp. 267–277.

[29] Doksum, K.A., and M. Gasko (1990): "On a correspondence between models in binary regression analysis and survival analysis," *International Statistical Review* 58, pp. 243–252.

[30] 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.

[31] 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.

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

[33] Firpo, S., N. Fortin, and T. Lemieux (2007): "Unconditional Quantile Regressions," *Econometrica,* forthcoming.

[34] 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.

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

[36] 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.

[37] Gutenbrunner, C., and J. Jurečková (1992): "Regression Quantile and Regression Rank Score Process in the Linear Model and Derived Statistics," *Annals of Statistics* 20, pp. 305-330.

[38] Hall, P., Wolff, R., and Yao, Q. (1999), "Methods for estimating a conditional distribution function," *Journal of the American Statistical Association* 94, pp. 154–163.

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

[40] 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.

[41] 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.

[42] 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.

[43] Imbens, G.W., and W. K. Newey (2009), "Identification and Estimation of Nonseparable Triangular Simultaneous Equations Models Without Additivity," *Econometrica,* forthcoming.

[44] Imbens, G.W., and J. Wooldridge (2008), "Recent Developments in the Econometrics of Program Evaluation," NBER Working Paper No. 14251.

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

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

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

[48] Lancaster, T. (1990): *The Econometric Analysis of Transition Data,* An Econometric Society Monograph, Cambridge University Press.

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

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

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

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

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

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

[55] 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.

[56] Rutemiller, H.C., and D.A. Bowers (1968): "Estimation in a Heteroscedastic Regression Model," *Journal of the American Statistical Association* 63, pp. 552-557.

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

[58] 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.

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

[60] 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, DR

		Effect of:			
Statistic	Total change	Minimum wage	Unions	Individual attributes	Coefficients
Men:					
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)	-27.1 (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)
Women:					
Standard	10.9 (0.4)	3.8 (0.1)	0.3 (0.0)	4.7 (0.2)	2.1 (0.3)
Deviation		34.9 (1.5)	3.2 (0.4)	42.8 (1.8)	19.1 (2.5)
90-10	39.8 (1.4)	23.0 (0.2)	0.9 (0.5)	14.5 (0.7)	1.3 (1.1)
		57.9 (1.9)	2.3 (1.2)	36.4 (1.7)	3.4 (2.6)
50-10	33.0 (0.7)	23.0 (0.2)	0.0 (0.1)	11.3 (0.4)	-1.4 (0.7)
		69.9 (1.6)	0.0 (0.4)	34.4 (1.3)	-4.3 (2.4)
90-50	6.8 (1.4)	0.0 (0.0)	0.9 (0.5)	3.1 (0.8)	2.8 (1.4)
		0.0 (0.0)	13.6 (7.2)	46.0 (11.3)	40.3 (9.9)
75-25	12.8 (0.9)	0.0 (0.0)	0.0 (0.5)	8.3 (0.2)	4.5 (0.8)
		0.0 (0.0)	0.0 (3.9)	65.1 (5.0)	35.0 (4.5)
95-5	38.8 (1.9)	16.8 (0.5)	0.7 (0.7)	16.4 (2.0)	5.0 (2.1)
		43.2 (2.2)	1.9 (1.9)	42.1 (5.0)	12.8 (5.1)
Gini	4.0 (0.1)	2.0 (0.1)	0.1 (0.0)	1.0 (0.1)	0.9 (0.1)
coefficient		49.0 (1.8)	3.5 (0.4)	24.5 (1.4)	23.0 (2.2)

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

Table 2: Decomposing Changes in Measures of Wage Dispersion: 1979-1988, CDR

		Effect of:			
Statistic	Total change	Minimum wage	Unions	Individual attributes	Coefficients
Men:					
Standard	8.2 (0.3)	3.3 (0.0)	0.6 (0.0)	1.9 (0.2)	2.4 (0.2)
Deviation		40.7 (1.4)	7.9 (0.5)	22.5 (1.8)	28.9 (2.4)
90-10	21.5 (1.0)	11.2 (0.1)	0.0 (0.0)	9.2 (0.8)	1.1 (1.3)
		52.1 (2.4)	0.0 (0.1)	42.6 (4.4)	5.3 (5.9)
50-10	11.3 (1.4)	11.2 (0.1)	-2.0 (1.0)	5.1 (0.4)	-3.1 (1.1)
		99.6 (14.1)	-17.9 (11.2)	45.5 (8.3)	-27.2 (14.0)
90-50	10.2 (1.2)	0.0 (0.0)	2.0 (1.0)	4.0 (0.8)	4.2 (1.1)
		0.0 (0.0)	19.7 (8.4)	39.3 (8.8)	41.0 (9.8)
75-25	15.4 (1.1)	0.0 (0.0)	4.1 (1.0)	0.3 (1.3)	11.1 (1.2)
		0.0 (0.0)	26.5 (6.2)	1.7 (8.6)	71.8 (8.7)
95-5	36.4 (2.1)	26.4 (0.7)	0.0 (0.6)	8.5 (1.1)	1.4 (1.5)
		72.7 (3.8)	0.0 (1.5)	23.4 (2.7)	3.9 (4.0)
Gini	4.2 (0.1)	1.6 (0.0)	0.4 (0.0)	0.3 (0.1)	1.8 (0.1)
coefficient		37.9 (1.1)	10.7 (0.5)	7.1 (1.6)	44.2 (1.6)
Women:					
Standard	12.7 (0.4)	5.6 (0.1)	0.3 (0.0)	5.1 (0.2)	1.7 (0.3)
Deviation		44.1 (1.5)	2.2 (0.4)	39.9 (1.8)	13.8 (2.5)
90-10	43.2 (1.4)	26.4 (0.2)	0.9 (0.5)	14.5 (0.7)	13.0 (1.1)
		61.2 (1.9)	2.2 (1.2)	33.5 (1.7)	3.1 (2.6)
50-10	36.4 (0.7)	26.4 (0.2)	0.0 (0.1)	11.3 (0.4)	-1.4 (0.7)
		72.7 (1.6)	0.0 (0.4)	31.2 (1.3)	-3.9 (2.4)
90-50	6.8 (1.4)	0.0 (0.0)	0.9 (0.5)	3.1 (0.8)	2.8 (1.4)
		0.0 (0.0)	13.6 (7.2)	46.0 (11.3)	40.3 (9.9)
75-25	12.8 (0.9)	0.0 (0.0)	0.0 (0.5)	8.3 (0.2)	4.5 (0.8)
		0.0 (0.0)	0.0 (3.9)	65.1 (5.0)	35.0 (4.5)
95-5	52.7 (1.9)	30.6 (0.5)	0.7 (0.7)	16.7 (2.0)	4.7 (2.1)
		58.1 (2.2)	1.4 (1.9)	31.6 (5.0)	8.8 (5.1)
Gini	4.9 (0.1)	2.9 (0.1)	0.1 (0.0)	1.3 (0.1)	0.6 (0.1)
coefficient		59.2 (1.8)	1.9 (0.4)	26.1 (1.4)	12.8 (2.2)

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 censored distribution regression model has been applied.

Table 3: Decomposing Changes in Measures of Wage Dispersion: 1979-1988, CQR

		Effect of:			
Statistic	Total change	Minimum wage	Unions	Individual attributes	Coefficients
Men:					
Standard	9.0 (0.3)	4.1 (0.0)	0.3 (0.0)	1.8 (0.1)	2.8 (0.2)
Deviation		45.3 (1.5)	3.2 (0.5)	20.0 (1.6)	31.5 (2.2)
90-10	22.3 (1.1)	14.2 (0.4)	-0.5 (0.1)	7.2 (0.4)	1.4 (1.1)
		63.6 (3.4)	-2.2 (0.6)	32.3 (2.8)	6.4 (5.1)
50-10	9.5 (0.9)	14.2 (0.4)	-1.8 (0.1)	4.6 (0.4)	-7.4 (0.9)
		148.7 (16.7)	-18.7 (3.0)	47.9 (9.0)	-78.0 (21.6)
90-50	12.7 (0.7)	0.0 (0.0)	1.3 (0.1)	2.6 (0.3)	8.8 (0.5)
		0.0 (0.0)	10.1 (1.0)	20.6 (2.4)	69.4 (2.5)
75-25	12.7 (0.6)	0.0 (0.0)	1.6 (0.1)	2.0 (0.4)	9.1 (0.5)
		0.0 (0.0)	12.9 (1.2)	15.5 (3.0)	71.5 (3.1)
95-5	39.2 (0.8)	30.6 (0.0)	-0.5 (0.1)	7.4 (0.5)	1.6 (0.8)
		78.1 (1.8)	-1.2 (0.3)	18.9 (1.2)	4.2 (2.1)
Gini	4.5 (0.1)	1.9 (0.0)	0.3 (0.0)	0.3 (0.1)	2.1 (0.1)
coefficient		42.2 (1.1)	5.9 (0.4)	6.1 (1.4)	45.9 (1.4)
Women:					
Standard	13.1 (0.4)	6.2 (0.0)	0.3 (0.1)	4.5 (0.3)	2.0 (0.3)
Deviation		47.6 (1.6)	2.6 (0.4)	34.8 (1.5)	15.0 (1.9)
90-10	48.8 (1.2)	30.6 (0.0)	0.8 (0.2)	14.7 (0.8)	2.7 (0.8)
		62.8 (1.5)	1.6 (0.3)	30.1 (1.3)	5.5 (1.6)
50-10	37.2 (0.7)	30.6 (0.0)	-0.3 (0.1)	10.9 (0.8)	-4.1 (0.5)
		82.3 (1.6)	-0.7 (0.3)	29.4 (1.7)	-10.9 (1.4)
90-50	11.5 (0.9)	0.0 (0.0)	1.1 (0.1)	3.7 (0.5)	6.7 (0.8)
		0.0 (0.0)	9.1 (1.1)	32.3 (3.3)	58.5 (3.5)
75-25	15.3 (0.9)	0.0 (0.0)	0.8 (0.1)	11.8 (0.8)	2.6 (0.9)
		0.0 (0.0)	5.6 (0.7)	77.6 (5.2)	16.9 (5.1)
95-5	50.8 (1.4)	30.6 (0.0)	1.1 (0.2)	15.1 (0.8)	4.0 (1.0)
		60.3 (1.6)	2.1 (0.4)	29.7 (1.2)	7.9 (1.8)
Gini	5.1 (0.1)	3.2 (0.0)	0.1 (0.0)	1.1 (0.1)	0.7 (0.1)
coefficient		62.5 (1.5)	2.1 (0.3)	21.8 (1.2)	13.6 (1.5)

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

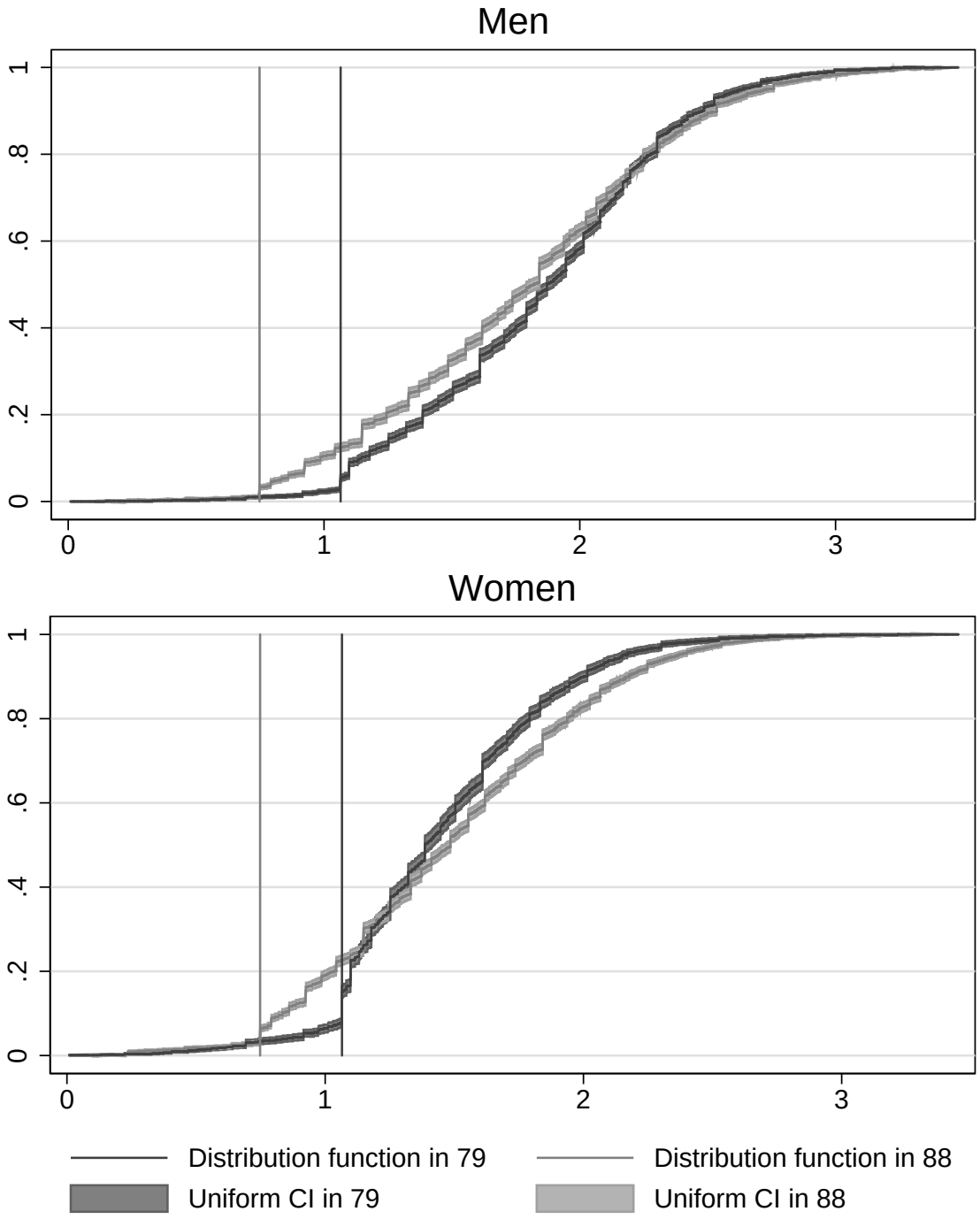

FIGURE 1. Empirical CDFs and 95% simultaneous confidence intervals for observed wages in 1979 and 1988. Distributions for men are plotted in the upper panel and distributions for women are plotted in the bottom panel. Confidence intervals were obtained by bootstrap with 100 repetitions. Vertical lines are the levels of the minimum wage.

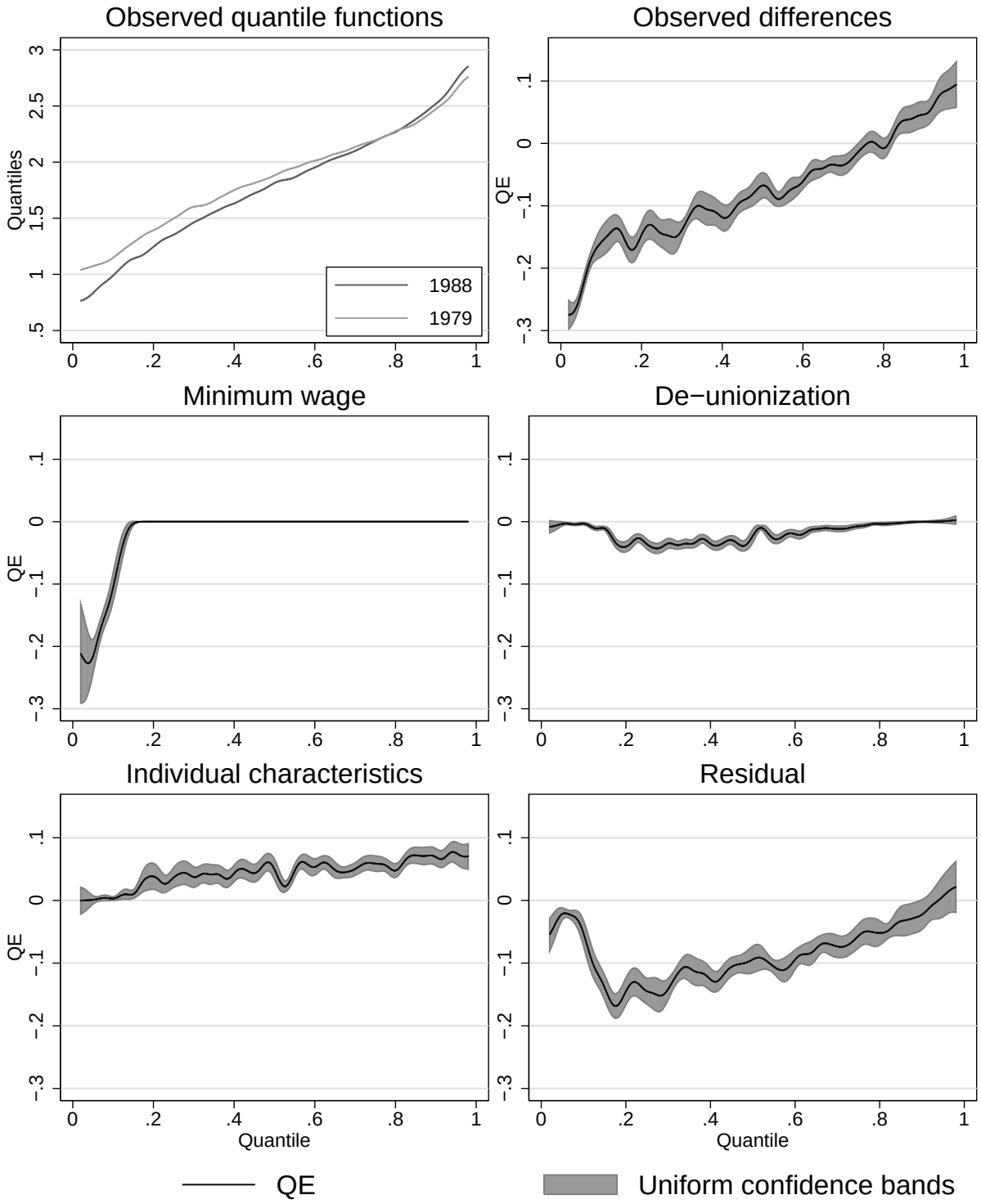

FIGURE 2. 95% simultaneous confidence intervals for observed quantile functions, observed quantile policy effects and decomposition of the quantile policy effects for men. Confidence intervals were obtained by bootstrap with 100 repetitions.

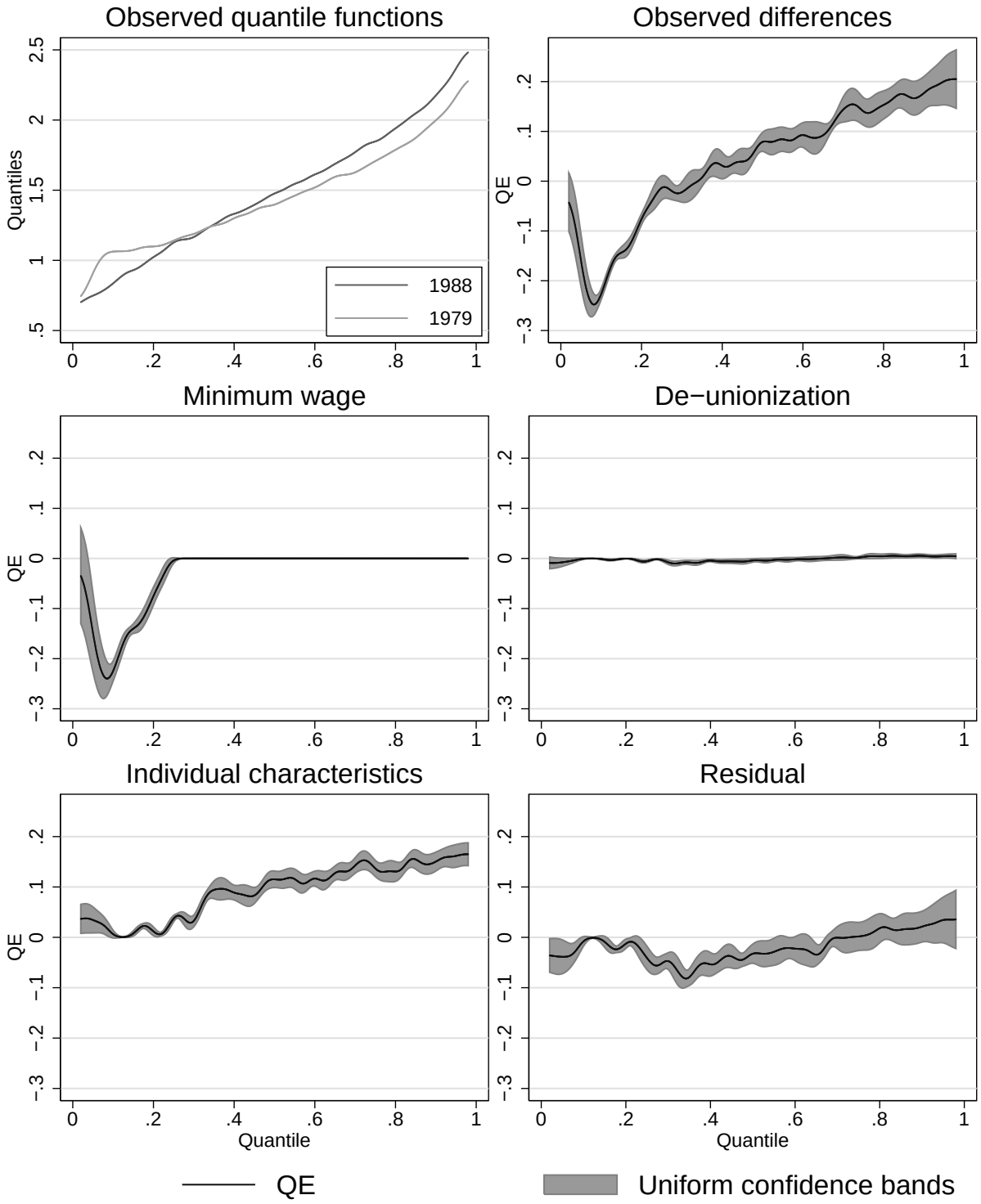

FIGURE 3. 95% simultaneous confidence intervals for observed quantile functions, observed quantile policy effects and decomposition of the quantile policy effects for women. Confidence intervals were obtained by bootstrap with 100 repetitions.

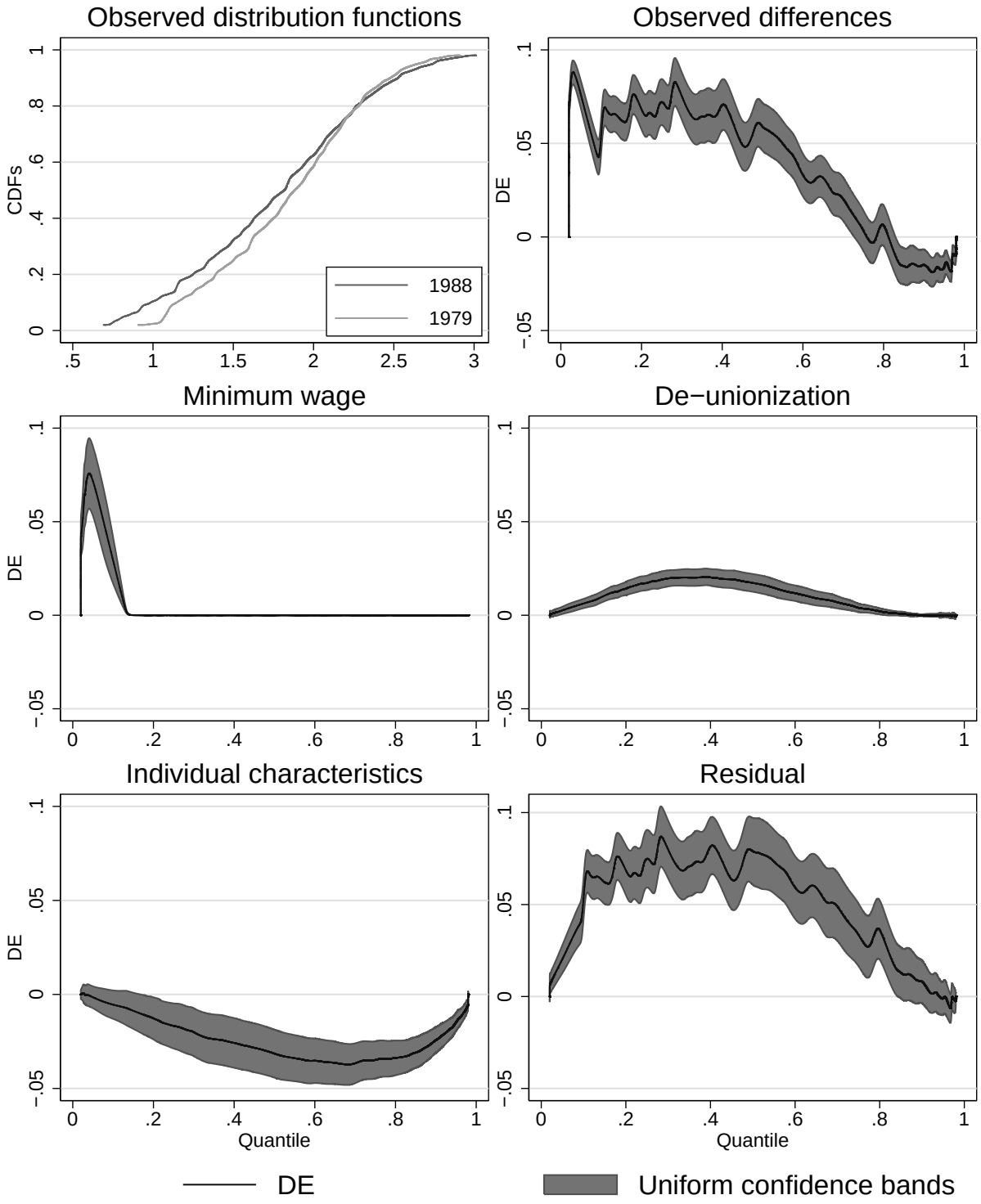

FIGURE 4. 95% simultaneous confidence intervals for observed distribution functions, observed distribution policy effects and decomposition of the distribution policy effects for men. Confidence intervals were obtained by bootstrap with 100 repetitions.

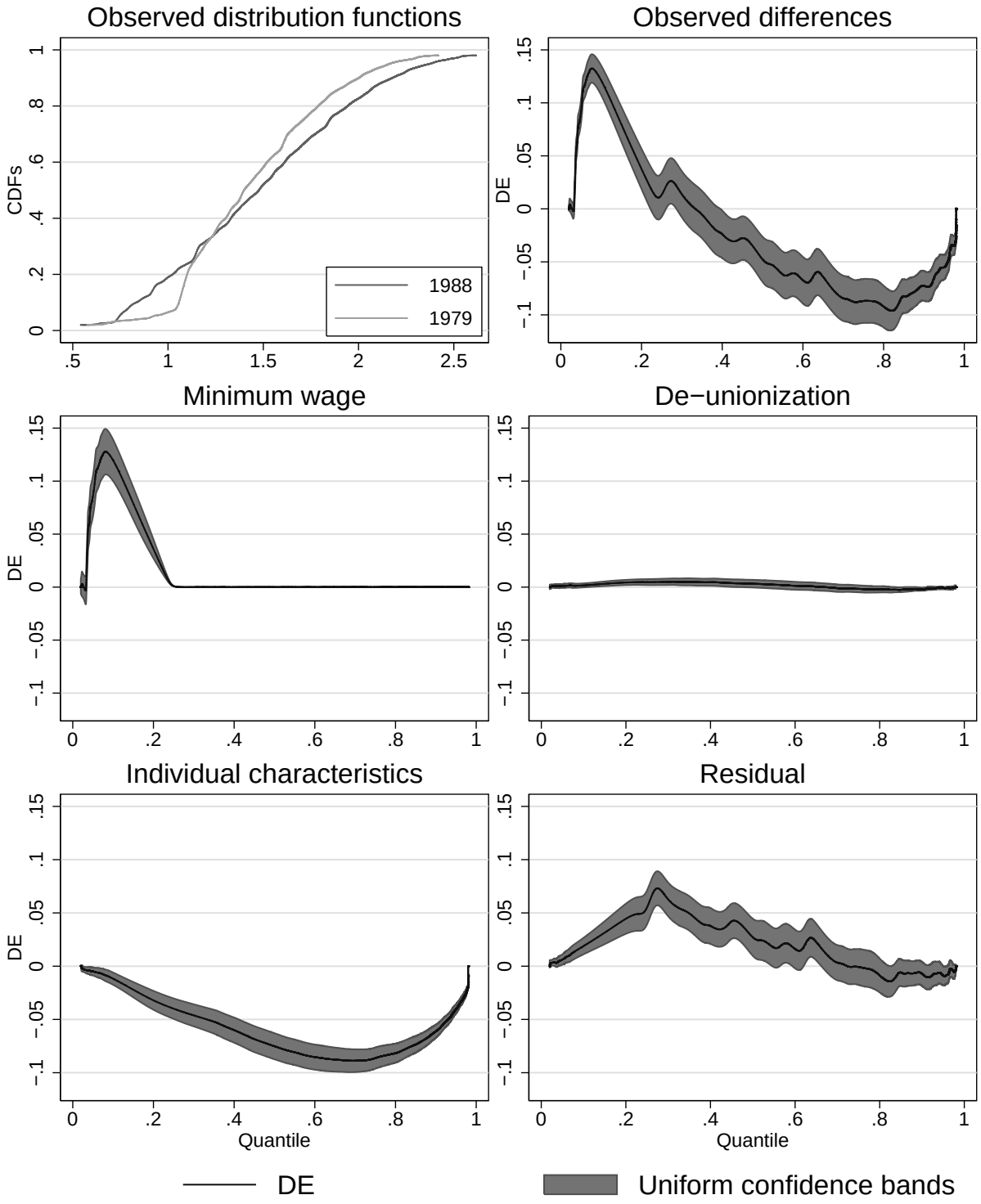

FIGURE 5. 95% simultaneous confidence intervals for observed distribution functions, observed distribution policy effects and decomposition of the distribution policy effects for women. Confidence intervals were obtained by bootstrap with 100 repetitions.

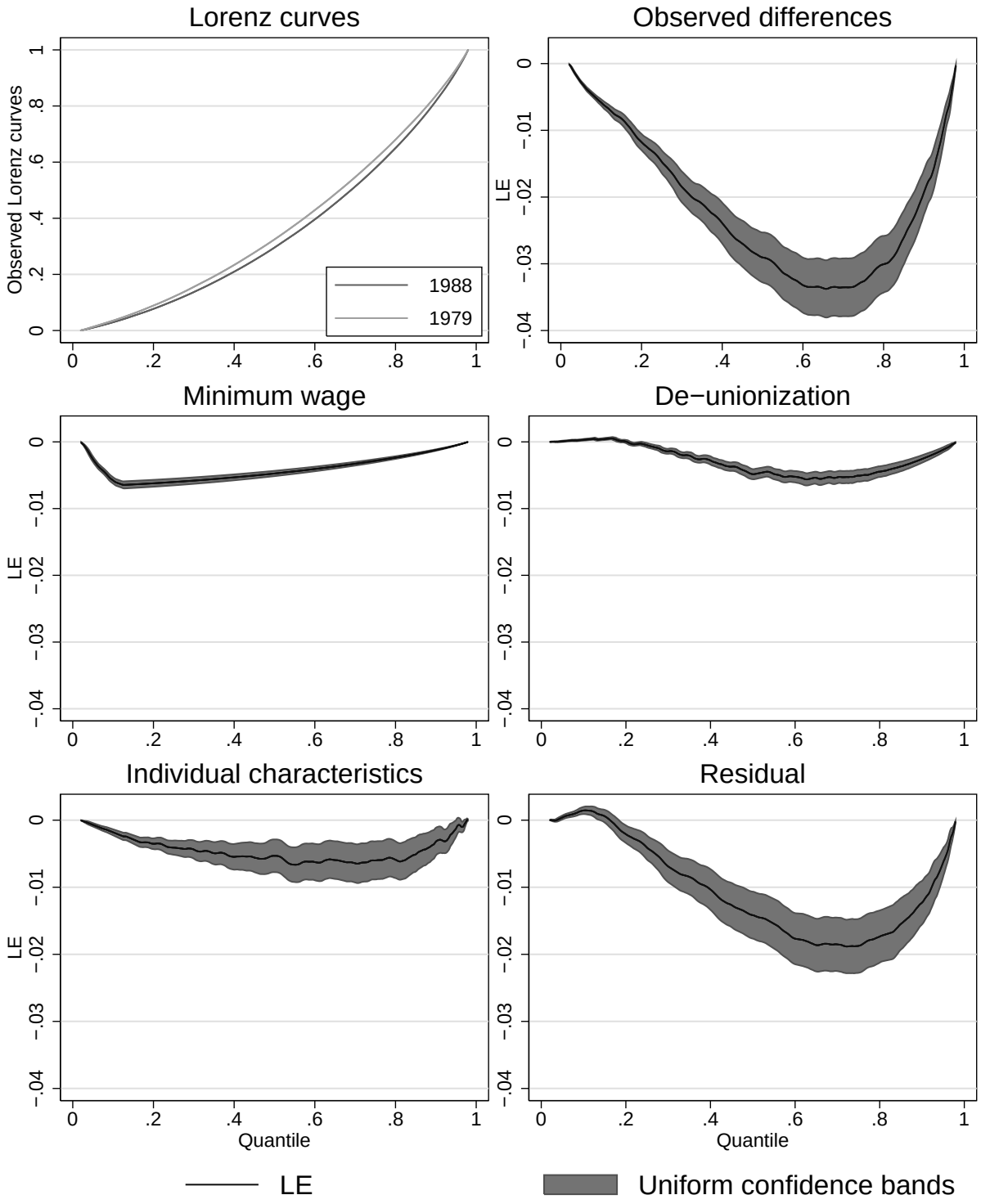

FIGURE 6. 95% simultaneous confidence intervals for observed Lorenz, observed Lorenz policy effects and decomposition of the Lorenz policy effects for men. Confidence intervals were obtained by bootstrap with 100 repetitions.

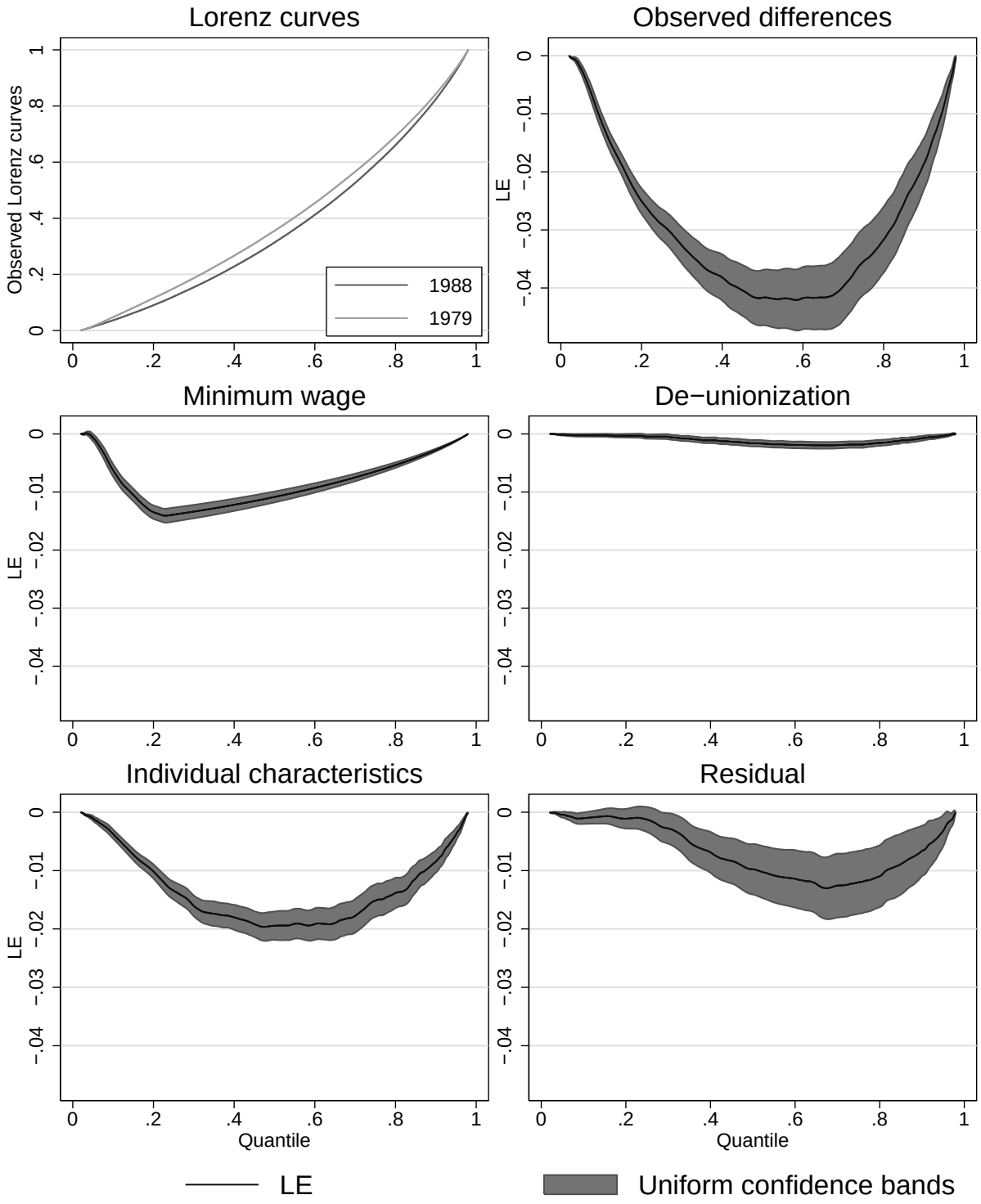

FIGURE 7. 95% simultaneous confidence intervals for observed Lorenz, observed Lorenz policy effects and decomposition of the Lorenz policy effects for women. Confidence intervals were obtained by bootstrap with 100 repetitions.

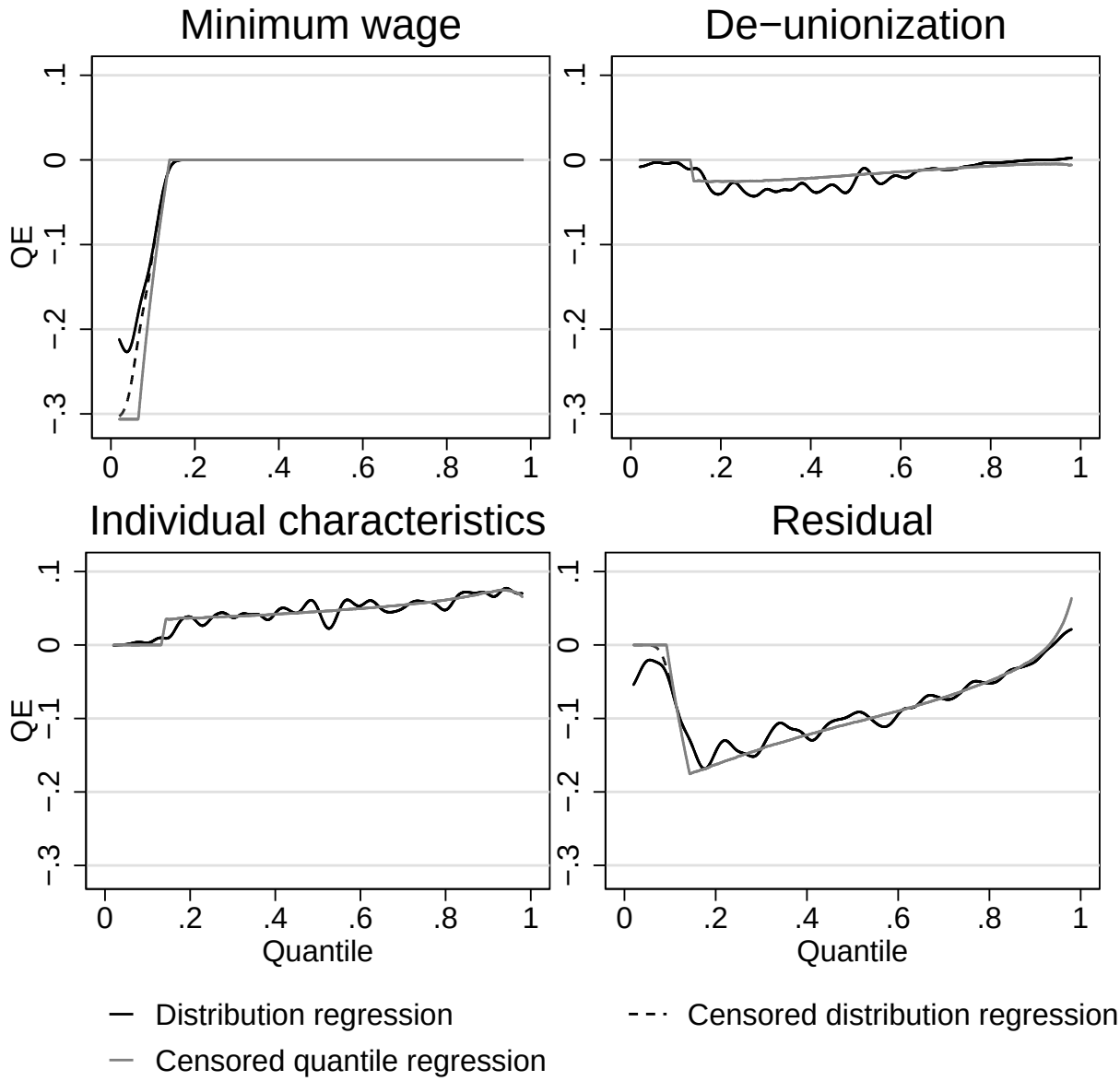

FIGURE 8. Comparison of distribution regression, censored distribution regression and censored quantile regression estimates of the decomposition of quantile policy effects for men.

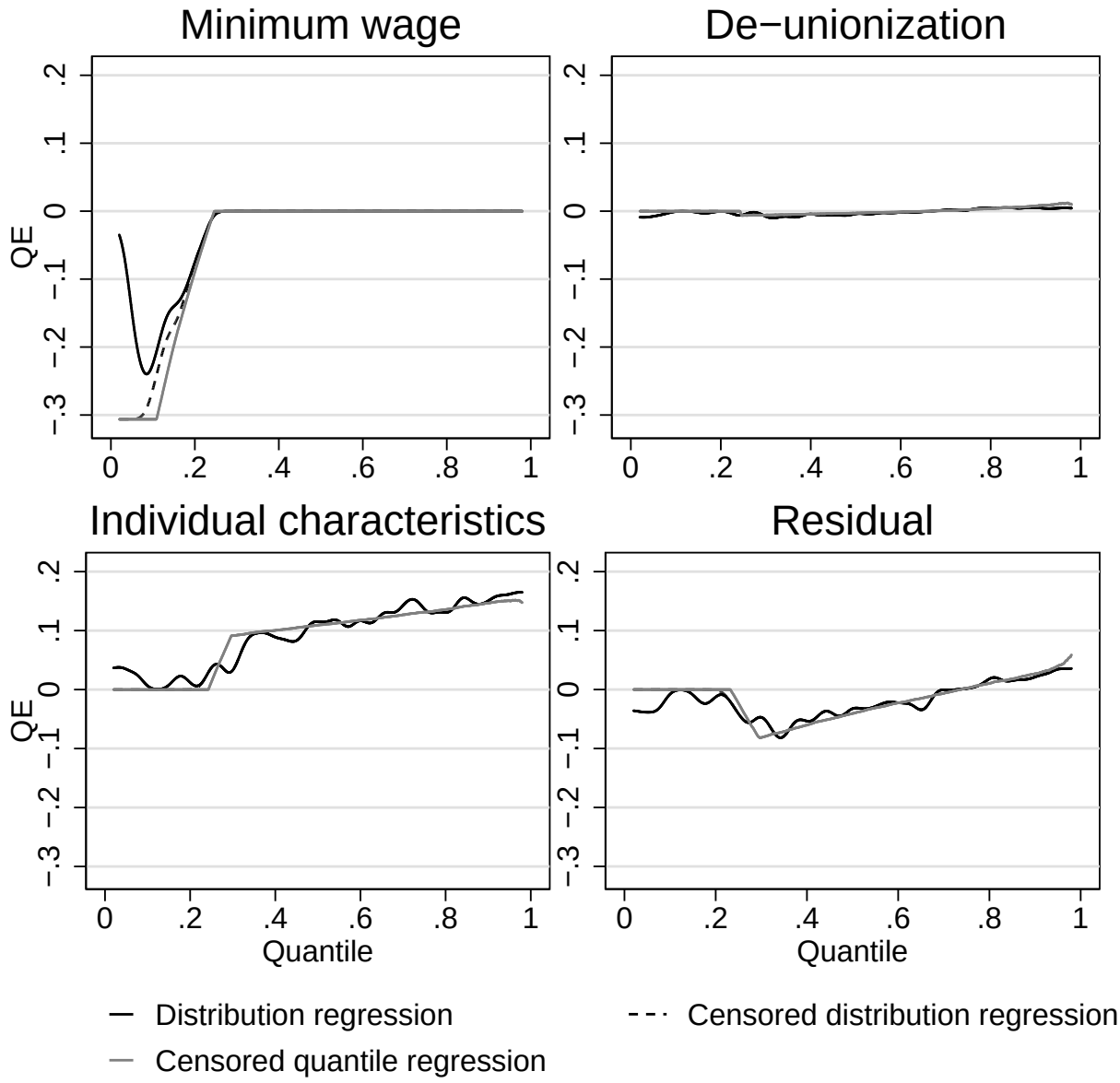

FIGURE 9. Comparison of distribution regression, censored distribution regression and censored quantile regression estimates of the decomposition of quantile policy effects for women.

Table A1: Reversing the order of the decomposition: 1979-1988, DR

		Effect of:			
Statistic	Total change	Individual attributes	Unions	Minimum wage	Coefficients
Men:					
Standard	8.0 (0.3)	0.9 (0.2)	1.5 (0.1)	2.9 (0.2)	2.7 (0.3)
Deviation		11.4 (2.8)	19.2 (1.0)	36.3 (2.8)	33.1 (2.4)
90-10	21.5 (1.0)	0.4 (1.3)	8.8 (1.2)	11.2 (0.9)	1.1 (1.3)
		1.8 (5.8)	40.7 (5.6)	52.1 (4.8)	5.3 (5.9)
50-10	11.3 (1.4)	2.5 (1.5)	0.7 (1.2)	11.2 (0.5)	-3.1 (1.1)
		21.8 (12.7	5.8 (11.3)	99.6 (15.5)	-27.1 (14.0)
90-50	10.2 (1.2)	-2.1 (1.2)	8.1 (0.8)	0.0 (0.9)	4.2 (1.1)
		-20.1 (12.9)	79.1 (10.7)	0.0 (9.5)	41.0 (9.8)
75-25	15.4 (1.1)	6.4 (1.2)	-2.1 (1.3)	0.0 (1.0)	11.1 (1.2)
		41.6 (6.8)	-13.4 (9.2)	0.0 (7.4)	71.8 (8.7)
95-5	33.0 (2.1)	2.6 (1.7)	5.9 (1.1)	23.0 (1.0)	1.4 (1.5)
		7.9 (4.8)	17.9 (3.4)	69.9 (5.5)	4.3 (4.4)
Gini	4.1 (0.1)	-0.3 (0.1)	1.0 (0.0)	1.4 (0.1)	2.0 (0.1)
coefficient		-6.8 (2.8)	24.1 (1.1)	33.3 (2.5)	49.4 (1.8)
Women:					
Standard	10.9 (0.4)	4.5 (0.2)	0.0 (0.0)	4.4 (0.3)	2.1 (0.3)
Deviation		41.1 (2.4)	-0.2 (0.2)	40.0 (2.4)	19.1 (2.5)
90-10	39.8 (1.4)	11.2 (0.7)	0.0 (0.4)	27.2 (0.2)	1.3 (1.1)
		28.2 (1.6)	0.0 (1.0)	68.4 (2.3)	3.4 (2.6)
50-10	33.0 (0.7)	7.9 (0.8)	-0.8 (0.5)	22.7 (0.5)	-1.4 (0.7)
		24.1 (2.6)	-2.4 (1.7)	82.6 (2.2)	-4.3 (2.4)
90-50	6.8 (1.4)	3.3 (0.8)	0.8 (0.5)	0.0 (0.5)	2.8 (1.4)
		47.9 (13.4)	11.7 (7.2)	0.0 (6.5)	40.3 (9.9)
75-25	12.8 (0.9)	2.8 (0.7)	0.0 (0.5)	5.5 (0.3)	4.5 (0.8)
		22.0 (5.6)	0.0 (3.9)	43.0 (3.5)	35.0 (4.5)
95-5	38.8 (1.9)	17.4 (0.9)	0.0 (0.3)	16.5 (2.0)	5.0 (2.1)
		44.8 (3.0)	0.0 (0.9)	42.4 (5.1)	12.8 (5.1)
Gini	4.0 (0.1)	0.6 (0.1)	0.0 (0.0)	2.5 (0.1)	0.9 (0.1)
coefficient		14.7 (2.4)	1.2 (0.3)	61.1 (2.7)	23.0 (2.2)

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 distribution regression model has been applied.