INFERENCE ON COUNTERFACTUAL DISTRIBUTIONS

VICTOR CHERNOZHUKOV† IVÁN FERNÁNDEZ-VAL§ BLAISE MELLY‡

Abstract. In this paper we develop procedures to make inference in regression models about how potential policy interventions affect the entire distribution of an outcome variable of interest. These policy interventions consist of counterfactual changes in the distribution of covariates related to the outcome. Under the assumption that the conditional distribution of the outcome is unaltered by the intervention, we obtain uniformly consistent estimates for functionals of the marginal distribution of the outcome before and after the policy intervention. Simultaneous confidence sets for these functionals are also constructed, which take into account the sampling variation in the estimation of the relationship between the outcome and covariates. This estimation can be based on several principal approaches for conditional quantile and distributions functions, including quantile regression and proportional hazard models. 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 of general form. An empirical application and a Monte Carlo example illustrate the results.

Date: May 21, 2008. Preliminary, comments are welcome. This paper replaces "Inference on Counterfactual Distributions Using Conditional Quantile Models," by V. Chernozhukov (first version of April 2005). We would like to thank Josh Angrist, Manuel Arellano, Flavio Cunha, and seminar participants at CEMFI, Columbia, MIT, and 2008 Winter Econometric Society Meetings for very useful comments that helped improve the paper.

† 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.

‡ Massachusetts Institute of Technology, Department of Economics. E-mail: melly@mit.edu.

1. Introduction

A common problem in economics is to predict the effect of a potential policy intervention or a counterfactual change in the economic conditions on some outcome variable of interest. For example, economists and policy analysts might be interested in what would have been the wage distribution in 2000 had the workers had the same characteristics as in 1990, what would have been the distribution of infant birth weights for black mothers had they received the same amount of prenatal care as white mothers, the effect on the distribution of food expenditure resulting from a change in income taxes, or the effect on the distribution of housing prices resulting from cleaning up a local hazardous-waste site. More generally, we can think of a policy intervention as a change in the distribution of a set of explanatory variables X that determine the response variable of interest Y. The policy analysis consists then in estimating the effect on the distribution of Y of a change in the distribution of X.

In this paper we develop procedures to make inference in regression models about how these counterfactual policy interventions affect the entire marginal distribution of Y. The main assumption is that the policy does not affect the relationship between the covariates and outcome. In other words, the conditional distribution of Y given X is not altered by the policy intervention. Starting from an estimate of the conditional model for the relationship between the outcome and covariates, we obtain uniformly consistent estimates for functionals of the distribution functions of the outcome before and after the intervention. Examples of these functionals include the own distribution functions, quantile functions, quantile treatment effects, distribution effects, means, variances, and Lorenz curves. Confidence sets are then constructed around the estimates that take into account the sampling variation coming from the estimation of the conditional model. These confidence sets are uniform in that they cover the entire functional of interest with pre-specified probability. The analysis is based upon several principal approaches to estimating conditional quantile functions and conditional distributions functions, including, for example, quantile regressions and proportional hazard models.

The proposed inference procedures 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 elaborated policies consisting of general changes in the covariate distribution. Moreover, the counterfactual distribution for the covariates can correspond to a known transformation of the values of the covariates in the population or to the covariate distribution in a different subpopulation or group. This variety of alternatives allows us to analyze, for instance, the effect of a redistribution of the covariates within the population or what would have been the counterfactual distribution of the outcome in one subpopulation had the covariates been distributed as in a different subpopulation.

To develop the statistical inference results, we establish the compact or Hadamard differentiability of the marginal distribution functions before and after the policy with respect to the limit of the estimators of the conditional model of the outcome given the covariates, tangentially to the set of continuous functions. 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 also facilitates to establish the validity of convenient resampling 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 covers such major methods as (1) conventional classical regression and its generalizations, (2) quantile regression and its generalization, (3) duration models, and (4) distribution regression models. As a consequence a wide array of techniques is covered;

in the discussion we devote most attention to the most practical and commonly used methods of estimating conditional quantities.

The results in the paper are related to the previous literature on policy estimators. Stocks (1989) develops nonparametric estimators to evaluate the mean effect of policy interventions. Gosling, Machin, and Meghir (2000) and Machado and Mata (2005) introduce policy estimators based on quantile regression models, but do not formally develop limit distribution theory for these estimators. This paper establishes the Gaussian limit distribution for the entire process of the counterfactual outcome distribution for a variety of policy estimators based on regression models, including location-scale models, conditional quantile models, proportional hazard models, and distribution regression models. To derive this result we formally establish the Hadamard differentiability of the counterfactual outcome distribution with respect to the limit of the conditional processes, which is required to apply the functional delta method. This functional differentiability approach is attractive from both a theoretical and practical point of view, because it allows us to establish the validity of resampling techniques for this problem, what greatly facilitates to carry out uniform inference about the entire distribution of policy effects. A recent paper by Firpo, Fortin, and Lemieux (2007) studies the effects of special policy interventions consisting of marginal changes in the values of the covariates. Their approach, based on a linearization of the functionals of interest, is clearly different from ours.

The rest of the paper is organized as follows. Section 2 describe methods to perform counterfactual analysis, setting up the modelling assumptions for the counterfactual outcomes and introducing the policy estimators. Section 3 derives limit distribution results for the policy estimators to perform uniform inference on functionals of the distribution of policy effects. Section 4 illustrates the estimation and inference procedures with numerical examples, and Section 5 concludes with a summary of the main results.

2. Methods for Counterfactual Analysis

2.1. **The model: 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 and are therefore observable, whereas counterfactual outcomes come from the population after the policy intervention and are therefore unobservable. We assume that the covariates are observable before and after the policy intervention. The observed outcomes are used to establish the relationship between the outcome and the covariates, which, together with the observed counterfactual distribution of the covariates, determine the distribution of the outcome after the intervention under some conditions that we make precise below.

For the purposes of specifying a model on how the counterfactual outcome is generated, it is convenient to look at the relationship between the observed outcome and covariates using a conditional quantile representation. Let Y^o be the observed outcome, and X^o be the $p \times 1$ vector of covariates with distribution function F_X^o before the policy intervention. Let $Q_Y(u|X)$ denote the conditional u-quantile of Y^o given X^o. The outcome Y^o can be linked to the conditional quantile function via the Skorohod representation:

$$Y^o = Q_Y(U^o|X^o), \text{ where } U^o \sim U(0,1) \text{ independently of } X^o \sim F_X^o. \tag{2.1}$$

This representation emphasizes that the outcome is a function of the covariates and the disturbance U^o. 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 both cases.

The counterfactual experiment consists of drawing the vector of covariates from a different distribution, i.e., $X^c \sim F_X^c$, where F_X^c is a known distribution function for the covariates after the policy intervention. Under the assumption that the conditional quantile function is not altered by the policy, the counterfactual outcome Y^c is generated by

$$Y^c = Q_Y(U^c|X^c), \text{ where } U^c \sim U(0,1) \text{ independently of } X^c \sim F_X^c. \quad (2.2)$$

Note that in the construction of the counterfactual outcome we make the additional assumption that the quantile function $Q_Y(u|x)$ can be evaluated at each point x in the support of the distribution of covariates F_X^c. This assumption either requires the support of F_X^c to be a subset of the support of F_X^o or that the quantile function can be suitably extrapolated outside the support of F_X^o.

The assumptions for the model that generates the counterfactual outcome can be stated formally as:

M.1 The conditional distribution of the outcome given the covariates is the same before and after the policy intervention.

M.2 The conditional model holds for all $x \in \mathcal{X}$, where $\mathcal{X}$ is a compact subset of $\mathbb{R}^p$ that contains the supports of F_X^o and F_X^c.

2.2. **Types of Counterfactual Changes.** We consider two different types of changes in the distribution of the covariates:

(1) The covariates are drawn from a different subpopulation before and after the intervention. These subpopulations might correspond to different demographic groups, time periods or geographic locations. Examples include the distributions of worker characteristic in different years, distributions of socioeconomic characteristics for black versus white mothers, or more generally distributions of covariates in a treatment group versus a control group.

(2) The policy intervention can be implemented as a known transformation of the distribution of the observed covariates; that is $X^c = g(X^o)$, where $g(\cdot)$ is a known function. This case covers, for example, unitary changes in the location of one of the covariates, $X^c = X + e_j$ where e_j is a unitary $p \times 1$ vector with a one in the position j; or mean preserving redistributions of the covariates implemented

as $X^c = (1-\alpha)E[X^o] + \alpha X^o$. This kind of policies can be used to estimate the effect on infant birth weights from an increase in the number of cigarettes smoked by the mother during pregnancy, the effect on food expenditure resulting from a change in income taxes, or the effect on housing prices resulting from cleaning up a local hazardous-waste site (Stock, 1991).

Note that these two cases correspond to conceptually different thought experiments. The statistical analysis that follows, however, will cover either situation without modification. The main difference will be that the second case corresponds to an almost perfectly controlled experiment, which provides additional information to identify more features of the joint distribution of the outcome before and after the intervention.

2.3. **Functionals of interest.** To make inference on the general effect on the outcome of the policy intervention, we need to identify the distribution and quantile functions of the outcome before and after the policy. The conditional distribution associated with the quantile function $Q_Y(u|x)$ is given by:

$$F_Y(y|x) = \int_0^1 1\{Q_Y(u|x) \leq y\}\, du. \tag{2.3}$$

Given our assumptions about how the counterfactual outcome is generated, the marginal distributions are given by

$$F_{Y^j}(y) := \Pr\left\{Y^j \leq y\right\} = \int_{\mathcal{X}} F_Y(y|x) dF_X^j(x), \tag{2.4}$$

with corresponding marginal u-quantile functions

$$Q_{Y^j}(u) = \inf\{y : F_{Y^j}(y) \geq u\}, \tag{2.5}$$

where j indexes the status before or after the policy, $j \in \{o, c\}$. The u-quantile treatment effect of the policy is then given by

$$QTE_Y(u) = Q_{Y^c}(u) - Q_{Y^o}(u). \tag{2.6}$$

Likewise, the y-distribution effect of the policy is given by

$$DE_Y(y) = F_{Y^c}(y) - F_{Y^o}(y). \tag{2.7}$$

Another functionals of interest might be the Lorenz curves of the observed and counterfactual outcomes. These curves, commonly used to measure inequality, are ratios of partial means to overall means

$$L_Y^j(y) := \int_{-\infty}^{y} t dF_Y^j(t) / \int_{-\infty}^{\infty} t dF_Y^j(t),$$

provided that the integrals exist and $\int_{-\infty}^{\infty} t dF_Y^j(t) \neq 0$, for $j \in \{o, c\}$. More generally, we might be interested in functionals of the marginal distributions of the outcome before and after the intervention

$$H_Y(y) := \phi(F_Y^o, F_Y^c, y). \tag{2.8}$$

These functionals include distributions, quantiles, quantile treatment effects, distribution effects, and Lorenz curves as special cases, but also other characteristics such as means with $\phi(F_Y^o, F_Y^c, y) = \int_{-\infty}^{\infty} t dF_Y^j(t) := \mu_Y^j$; mean effects with $\phi(F_Y^o, F_Y^c, y) = \mu_Y^c - \mu_Y^o$; variances with $\phi(F_Y^o, F_Y^c, y) = \int_{-\infty}^{\infty} t^2 dF_Y^j(t) - (\mu_Y^j)^2 := (\sigma_Y^j)^2$; and variance effects with $\phi(F_Y^o, F_Y^c, y) = (\sigma_Y^c)^2 - (\sigma_Y^o)^2$.[1]

In the case where the policy consists of a known transformation of the distribution of the covariates, $X^c = g(X^o)$, we can also identify the distribution and quantile functions for the effects of the policy, $\Delta = Y^c - Y^o$, by:

$$F_\Delta(\delta) = \int_{\mathcal{X}} \int_0^1 1\left\{Q_Y(u|g(x)) - Q_Y(u|x) \leq \delta\right\} du dF_X^o(x) \tag{2.9}$$

and

$$Q_\Delta(u) = \inf\{\delta : F_\Delta(\delta) \geq u\}, \tag{2.10}$$

[1] In the rest of the discussion we keep the distribution, quantile, quantile treatment effects, and distribution effects functions as separate cases to emphasize the importance of these functionals in practice. Lorenz curves are special cases of the general functional with $\phi(F_Y^o, F_Y^c, y) = \int_{-\infty}^{y} t dF_Y^j(t) / \int_{-\infty}^{\infty} t dF_Y^j(t)$, and will not be considered separately.

under the additional assumption

RP Conditional rank preservation: $U^c = U^o | X^o$.

See, e.g., Heckman, Smith, and Clements (1997).

2.4. **Conditional models.** The quantile representation for the relationship between outcome and covariates is a useful modeling tool to understand how the counterfactual outcomes are generated, but it is not necessary for identifying the marginal distribution and quantile functions. These functions depend only on the conditional distribution of the outcome, so we can proceed either by directly specifying a model for the conditional distribution, or by specifying a model for the conditional quantiles and then obtaining the conditional distribution using the expression (2.3). We next describe several principal methods for modeling and estimating conditional quantile and distribution functions.

Example 1. Location regression and generalizations. The inference results of this paper cover the classical regression model as well as its generalizations. The classical location-shift model takes the form

$$Y = m(X) + V, \quad V = Q_V(U), \tag{2.11}$$

where $U \sim U(0,1)$ is independent of X, and $m(\cdot)$ is a location functional, for example, 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 can have heterogeneous effects on the entire marginal distribution of Y, affecting its various quantiles in a differential manner. The regression function $m(x)$ is most commonly modeled linearly in parameters $m(x) = x'\beta$ and estimated using least squares or instrumental variable methods. The quantile function $Q_V(u)$ can be left unrestricted and estimated 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 central limit theorem.

The location model has played a classical role in the regression model. Most endogenous and exogenous treatment effects models, for example, can be analyzed and estimated using variations of this model; see, e.g., Chap. 25 in Cameron and Trivedi (2005). A variety of standard survival and duration models also imply (2.11) after a transformation, e.g., the Cox models with Weibull hazards and accelerated failure time models, cf. Docksum and Gasko (1990).

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

$$Y = m(X) + \sigma(X) \cdot V, \quad V = Q_V(U), \tag{2.12}$$

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 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, for example, Rutemiller and Bowers (1968) and Koenker and Xiao (2002).

Example 2. Quantile regression. The quantile regression method directly models the conditional quantile relationship

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

without imposing the location shift or the location-scale shift mechanisms. The model permits the covariates to impact Y by changing not only the location and scale of the distribution but also the entire shape. An early convincing example of such effects goes back to Doksum (1974), who showed that regression data can be sharply inconsistent with the location-scale shift paradigm. Quantile regression precisely addresses this issue.

The leading approach to quantile regression entails the approximation of the conditional quantile function by a linear functional form $Q_Y(u|x) = x'\beta(u)$, see, e.g., Koenker and Bassett (1978) and Koenker (2005).[2]

Example 3. Duration Models. A common way to model the distribution functions in duration and survival analysis is the Cox model:

$$F_Y(y|x) = \exp(\exp(m(x) + t(y))),$$

where $t(y)$ is a monotonic function in y. 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 regression representation:

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

where V has an extreme value distribution. Therefore covariates impact the outcome only through the location function. The estimation of this model has been a subject of many studies, e.g., Lancaster (1990) and Dabrowska (2005).

Example 4. Distribution Regression. Instead of restricting ourselves to the model of the above kind, 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 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 (see, e.g., Foresi and Peracchi, 1995). This approach is similar to quantile regression in spirit. In particular, as quantile regression, this approach leads

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

to the specification $Y = Q_Y(U|X) = \Lambda^{-1}(m^{-1}(U,X))$ where $U \sim U(0,1)$ independently of X.

2.5. **Policy estimators.** Given an estimator for the conditional distribution $\hat{F}_Y(y|x)$, the marginal distributions for the outcome can be estimated by

$$\hat{F}^j_Y(y) = \int \hat{F}_Y(y|x) dF^j_X(x), \tag{2.13}$$

with corresponding quantile functions

$$\hat{Q}^j_Y(u) = \inf\{y : \hat{F}^j_Y(y) \geq u\}, \tag{2.14}$$

for $j \in \{o, c\}$. Estimators for the quantile treatment effects and distribution effects can then be constructed as

$$\widehat{QTE}_Y(u) = \hat{Q}^c_Y(u) - \hat{Q}^o_Y(u), \tag{2.15}$$

and

$$\widehat{DE}_Y(y) = \hat{F}^c_Y(y) - \hat{F}^o_Y(y). \tag{2.16}$$

For the general functionals introduced in (2.8), we can use sample analogs

$$\hat{H}_Y(y) = \phi(\hat{F}^o_Y, \hat{F}^c_Y, y). \tag{2.17}$$

In the previous expressions the estimator for the conditional distribution can be obtained directly from a conditional distribution model, or by inversion of a conditional quantile estimator, that is:

$$\hat{F}_Y(y|x) = \int_0^1 1\left\{\hat{Q}_Y(u|x) \leq y\right\} du, \tag{2.18}$$

where $\hat{Q}_Y(u|x)$ is a given estimator of the conditional quantile function and the integral can be approximated by a sum over a fine grid of the interval $[0,1]$. Estimators for the distribution and quantiles of the effects can be constructed similarly by replacing the conditional functions by their estimators in expressions (2.9) and (2.10).

2.6. **Inference questions.** Common inference questions that arise in policy analysis involve features of the distribution of the outcome variable before and affect the intervention. For example, we might be interested in the average effect of the policy, or in quantile treatment effects at several quantiles that measure the impact of the policy at different parts of the outcome distribution. More generally, in this analysis is very common that the questions of interest involve the entire distribution or quantile functions of the outcomes. Examples include the hypotheses that the policy has no effect, that the effect is constant, or that it is positive for the entire distribution. 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. The next section provides limit distribution theory for the policy estimators. This theory applies to the entire distribution and quantile functions of the outcome before and after the response, and therefore is valid to make both pointwise inference about specific features of these functions, and simultaneous inference about the entire distribution function, quantile function, or other functionals of interest.

3. Limit Distribution Theory for Policy Estimators

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

3.1. **Estimators of the Conditional Model.** We provide general assumptions about the estimators of the conditional model for the relationship between the outcome and covariates, which will allow us to derive the limit distribution for the policy estimators constructed from them. These assumptions hold for commonly used parametric and semiparametric estimation methods for conditional distribution and quantile models such as linear quantile regression and proportional hazard models. We provide separate assumptions for quantile and distribution estimators, and then show that in both cases the ultimate estimator of the conditional distribution used to obtain the functionals of interest satisfy a functional central limit theorem, what enables us to give a unified treatment for all the policy estimators.

We start the analysis by discussing the approach based on quantile models. By $\ell^\infty((0,1)\times\mathcal{X})$ we denote the space of bounded functions mapping from $(0,1)\times\mathcal{X}$ to $\mathbb{R}$, equipped with the uniform metric. The following conditions impose some restrictions on the conditional quantile model and on the corresponding quantile estimator:

C.1 There exists a conditional density $f_Y(y|x)$ that 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 $\mathbb{R}$.

Q.1 The estimator of the conditional quantile function $(u,x) \mapsto \hat{Q}_Y(u|x)$ converges in law to a continuous Gaussian process:

$$\sqrt{n}\left(\hat{Q}_Y(u|x) - Q_Y(u|x)\right) \Rightarrow V(u,x), \tag{3.1}$$

in the space $\ell^\infty((0,1)\times\mathcal{X})$, where the random function $(u,x)\mapsto V(u,x)$ has zero mean and uniformly bounded covariance function $\Sigma_V(u,x,\tilde{u},\tilde{x}) := E[V(u,x)V(\tilde{u},\tilde{x})]$.

These conditions appear reasonable in practice when the outcome is continuous. If the outcome is discrete the condition **C.1** does not hold in the stated form. However, we can deal with discrete outcomes via the distribution approach. Condition **C.1** focuses on the case where the outcome has compact support with bounded density, and is a reasonable case to analyze in detail first. This condition could be extended to include other cases, through none of the subsequent results are expected to change in an essential manner. Condition **C.2** applies to the main estimators of conditional quantile functions under suitable regularity conditions, cf., Gutenbrunner and Jureckova (1992) and Angrist, Chernozhukov, and Fernandez-Val (2006).

Turning to the conditional distribution estimators, 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 $\mathbb{R}$. The following condition imposes some regularity conditions on the way the estimator of the distribution function should behave.

D.1 The estimated conditional distribution function $(u, x) \mapsto \hat{F}_Y(y|x)$ converges in law to a continuous Gaussian process:

$$\sqrt{n}\left(\hat{F}_Y(y|x) - F_Y(y|x)\right) \Rightarrow Z(y, x), \tag{3.2}$$

in the space $\ell^\infty(\mathcal{Y} \times \mathcal{X})$, where the random function $(y, x) \mapsto Z(y, x)$ has zero mean and uniformly bounded covariance function $\Sigma_Z(y, x, \tilde{y}, \tilde{x}) := E[Z(y, x)Z(\tilde{y}, \tilde{x})]$.

This condition holds for common estimators of conditional distribution functions, see, e.g., Beran (1977), Owen (1987), Andrews (1988) and Andrews (1993). These estimators, however, might produce estimates that are not monotonic in the level of the outcome y, see, e.g., 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, see Chernozhukov, Fernandez-Val, and Galichon (2006). Start from an estimator $\tilde{F}_Y(y|x)$ that satisfies condition **D.1**, but it is not necessarily monotonic in y. This estimator can be rearranged with the following two steps. First, construct

$$\hat{Q}(u|x) = \int_0^\infty 1\{\tilde{F}_Y(y|x) < u\}dy - \int_{-\infty}^0 1\{\tilde{F}_Y(y|x) > u\}dy, \tag{3.3}$$

which is an estimator of the conditional quantile function that is monotone in the quantile index u. Second, invert $\hat{Q}(u|x)$ to obtain

$$\hat{F}_Y(y|x) = \inf\{u : \hat{Q}(u|x) \leq y\}, \tag{3.4}$$

a monotone estimator of the conditional conditional function, which, under the assumption **C.1**, has the same first order limit distribution as $\tilde{F}_Y(y|x)$, see Chernozhukov, Fernandez-Val, and Galichon (2006).

If we start from a conditional quantile model, we can use the relationship between the distribution function and the quantile function to define the conditional distribution function estimator in (2.18) from an available conditional quantile function estimator $\hat{Q}_Y(u|x)$. It turns out that if the original quantile estimator satisfies the conditions **C.1** and **Q.1**, then the implied conditional distribution estimator satisfies the condition **D.1**. This result is convenient for the following analysis because it allows us to give a unified treatment of the policy estimators based on both quantile models and distribution models.

Lemma 1. *Under the conditions **C.1** and **Q.1**, the estimator of the conditional distribution function in (2.18) satisfies the condition **D.1** with*

$$Z(y, x) = -f_Y(y|x)V(F_Y(y|x), x). \tag{3.5}$$

3.2. **Basic principles.** The derivation of the limit distribution for the policy estimators is based on two basic principles that allow us to link the properties of the conditional estimators with the properties of the estimators of the marginal distribution and quantile functions. First, although there does not exist a direct connection between conditional and marginal quantiles, the law of iterated expectations links conditional and marginal distributions. Second, by means of delta method we can switch from the properties of the estimators of the distribution function to the properties of the estimators of the quantile function and vice versa. The main difficulty in the analysis is that the functionals of interest depend on the entire process for the conditional function, so we need to resort to a functional delta method. Moreover, the estimators of the conditional model usually have discontinuities because their estimating equations involve indicator functions, what further complicates the analysis.

The key ingredient in the derivation and the main theoretical contribution of the paper is to show the Hadamard or compact differentiability of the functionals of interest with respect to the limit of the conditional processes, tangentially to the subspace of continuous functions. The basic Hadamard differentiability result for the conditional distribution with respect to the conditional quantile function is given in Lemma 4 in the Appendix, and the differentiability of the other functionals then follows by the properties of the Hadamard derivative. These results enable us to use the functional delta method to derive all the following limit distribution theory.

3.3. **Limit distribution for marginal distribution and quantile functions.** We are now ready to state the first main results establishing that the estimators of the marginal distribution and quantile functions satisfy a central limit theorem in large samples.

Theorem 1 (Limit Distribution for Marginal Distributions)**.** *Under conditions **M.1**, **M.2**, and **D.1** the estimators of the marginal distribution functions converge in law to the following continuous linear functional of the Gaussian process $Z(y,x)$:*

$$\sqrt{n}\left(\hat{F}_Y^j(y) - F_Y^j(y)\right) \Rightarrow \int_{\mathcal{X}} Z(y,x)dF_X^j(x) := Z^j(y), \tag{3.6}$$

in the space $\ell^\infty(\mathcal{Y})$, where the random function $y \mapsto Z^j(y)$ has zero mean and covariance function

$$\Sigma_Z^j(y,\tilde{y}) := \int_{\mathcal{X}}\int_{\mathcal{X}} \Sigma_Z(y,x,\tilde{y},\tilde{x})dF_X^j(x)dF_X^j(\tilde{x}). \tag{3.7}$$

The convergence holds jointly for all estimators indexed by the status $j \in \{o,c\}$, with cross covariance function

$$\Sigma_Z^{oc}(y,\tilde{y}) := \int_{\mathcal{X}}\int_{\mathcal{X}} \Sigma_Z(y,x,\tilde{y},\tilde{x})dF_X^o(x)dF_X^c(\tilde{x}). \tag{3.8}$$

Theorem 2 (Limit Distribution for Marginal Quantiles)**.** *Under the conditions **M.1**, **M.2**, **C.1**, and **D.1** the estimators of the marginal quantile functions converge in law to the following continuous linear Gaussian functional:*

$$\sqrt{n}\left(\hat{Q}_Y^j(u) - Q_Y^j(u)\right) \Rightarrow -f_Y^j(Q_Y^j(u))^{-1}Z^j(Q_Y^j(u)) := V^j(u), \tag{3.9}$$

in the space $\ell^\infty((0,1))$, where $f_Y^j(y) = \int_{\mathcal{X}} f_Y(y|x)dF_X^j(x)$ and the random function $u \mapsto V^j(u)$ has zero mean and covariance function

$$\Sigma_V^j(u,\tilde{u}) := f_Y^j(Q_Y^j(u))^{-1}f_Y^j(Q_Y^j(\tilde{u}))^{-1}\Sigma_Z^j(Q_Y^j(u),Q_Y^j(\tilde{u})). \tag{3.10}$$

The convergence holds jointly for all estimators indexed by the status $j \in \{o, c\}$, with cross-covariance function

$$\Sigma_V^{oc}(u, \tilde{u}) := f_Y^o(Q_Y^o(u))^{-1} f_Y^c(Q_Y^c(\tilde{u}))^{-1} \Sigma_Z^{oc}(Q_Y^o(u), Q_Y^c(\tilde{u})). \tag{3.11}$$

Corollary 1 (Limit Distribution for Quantile Treatment Effects)**.** *Under the conditions **M.1**, **M.2**, **C.1**, and **D.1** the estimator of the quantile treatment effects converges in law to the following linear functional of continuous Gaussian processes:*

$$\sqrt{n}\left(\widehat{QTE}_Y(u) - QTE_Y(u)\right) \Rightarrow V^c(u) - V^o(u) := W(u), \tag{3.12}$$

in the space $\ell^\infty((0,1))$, where the random function $u \mapsto W(u)$ has zero mean and covariance function

$$\Sigma_W(u, \tilde{u}) := \Sigma_V^o(u, \tilde{u}) + \Sigma_V^c(u, \tilde{u}) - \Sigma_V^{oc}(u, \tilde{u}) - \Sigma_V^{oc}(\tilde{u}, u). \tag{3.13}$$

Corollary 2 (Limit Distribution for Distribution Effects)**.** *Under the conditions **M.1**, **M.2**, and **D.1** the estimator of the distribution effects converges in law to the following linear functional of continuous Gaussian processes:*

$$\sqrt{n}\left(\widehat{DE}_Y(y) - DE_Y(y)\right) \Rightarrow Z^c(u) - Z^o(u) := S(y), \tag{3.14}$$

in the space $\ell^\infty(\mathcal{Y})$, where the random function $y \mapsto S(y)$ has zero mean and covariance function

$$\Sigma_S(y, \tilde{y}) := \Sigma_Z^o(y, \tilde{y}) + \Sigma_Z^c(y, \tilde{y}) - \Sigma_Z^{oc}(y, \tilde{y}) - \Sigma_Z^{oc}(\tilde{y}, y). \tag{3.15}$$

Corollary 3 (Limit Distribution for Differentiable Functionals)**.** *Let $H_Y(y) = \phi(F_Y^o, F_Y^c, y)$ be a Hadamard differentiable functional in the first two arguments, with derivatives ϕ_o' and ϕ_c' with respect to the first and second argument. Under the conditions **M.1**, **M.2**, and **D.1** the estimator of the functional $H_Y(y)$ defined in (2.17) converges in law to the following linear functional of continuous Gaussian processes:*

$$\sqrt{n}\left(\hat{H}_Y(y) - H_Y(y)\right) \Rightarrow \phi_o'(F_Y^o, F_Y^c, y) Z^o(y) + \phi_c'(F_Y^o, F_Y^c, y) Z^c(y) := T(y), \tag{3.16}$$

in the space $\ell^\infty(\mathcal{Y})$, where the random function $y \mapsto T(y)$ has zero mean and covariance function

$$\Sigma_T(y,\tilde{y}) := \phi'_o\tilde{\phi}'_o\Sigma^o_Z(y,\tilde{y}) + \phi'_c\tilde{\phi}'_c\Sigma^c_Z(y,\tilde{y}) + \phi'_o\tilde{\phi}'_c\Sigma^{oc}_Z(y,\tilde{y}) + \tilde{\phi}'_o\phi'_c\Sigma^{oc}_Z(\tilde{y},y), \tag{3.17}$$

where $\phi'_j := \phi'_j(F^o_Y, F^c_Y, y)$ and $\tilde{\phi}'_j := \phi'_j(F^o_Y, F^c_Y, \tilde{y})$, for $j \in \{o, c\}$.

Remark 1. *The previous Corollary follows from the functional delta method; see, e.g., Theorem 20.8 in van der Vaart (1998). Examples of Hadamard differentiable functionals include continuous transformations of linear functionals such as means, mean effects, variances, variance effects, covariances, and Lorenz curves; cf. Fernholz (1983).*

3.4. **Limit distribution for the estimators of the effects.** For policy interventions that can be implemented as a known transformation of the covariate, $X^c = g(X^o)$, we can also identify and estimate the distribution of effects under the additional assumption of conditional rank preservation. The following results provide estimators for the distribution and quantile functions of the effects and limit distribution theory for them.

Lemma 2 (Limit distribution for estimators of conditional distribution and quantile functions)**.** *Let $\hat{Q}_\Delta(u|x) = \hat{Q}_Y(u|g(x)) - \hat{Q}_Y(u|x)$ be an estimator of the conditional quantile function of the effects $Q_\Delta(u|x)$.*[3] *Under the conditions **C.1**, **Q.1**, and **R.P**, we have:*

$$\sqrt{n}\left(\hat{Q}_\Delta(u|x) - Q_\Delta(u|x)\right) \Rightarrow V(u, g(x)) - V(u,x) := V_g(u,x), \tag{3.18}$$

in the space $\ell^\infty((0,1)\times\mathcal{X})$, where the Gaussian random function $(u,x) \mapsto V_g(u,x)$ has zero mean and covariance function

$$\Omega_V(u,x,\tilde{u},\tilde{x}) := \Sigma^o_V(u,g(x),\tilde{u},g(\tilde{x})) + \Sigma^o_V(u,x,\tilde{u},\tilde{x}) - 2\Sigma^o_V(u,g(x),\tilde{u},\tilde{x}). \tag{3.19}$$

[3] In the distribution approach, $\hat{Q}_Y(u|x)$ can be obtained by inversion of the estimator of the conditional distribution as in (3.3).

Let $\hat{F}_\Delta(\delta|x) = \int_0^1 \mathbf{1}\{\hat{Q}_\Delta(u|x) \leq \delta\}du$ *be an estimator of the conditional distribution of the effects* $F_\Delta(\delta|x)$*. Under the conditions* ***C.1****,* ***Q.1****, and* ***R.P****, we have:*

$$\sqrt{n}\left(\hat{F}_\Delta(\delta|x) - F_\Delta(\delta|x)\right) \Rightarrow -f_\Delta(\delta|x)V_g(F_\Delta(\delta|x), x) := Z_g(\delta, x), \tag{3.20}$$

in the space $\ell^\infty(\mathcal{D} \times \mathcal{X})$*, where* $\mathcal{D} = \{\delta \in \mathbb{R} : \delta = y - \tilde{y}, y \in \mathcal{Y}, \tilde{y} \in \mathcal{Y}\}$*, and the random function* $(\delta, x) \mapsto Z_g(\delta, x)$ *has zero mean and covariance function*

$$\Omega_Z(\delta, x, \tilde{\delta}, \tilde{x}) := f_\Delta(\delta|x)f_\Delta(\tilde{\delta}|\tilde{x})\Omega_V(F_\Delta(\delta|x), x, F_\Delta(\tilde{\delta}|\tilde{x}), \tilde{x}). \tag{3.21}$$

The conditional density of the effect, $f_\Delta(\delta|x)$*, assumed to be bounded above and away from zero,*[4] *can be expressed in terms of the conditional density of the level of the outcome as*

$$f_\Delta(\delta|x) = \left[\frac{1}{f_Y\left(Q_Y(F_\Delta(\delta|x)|g(x))|g(x)\right)} - \frac{1}{f_Y\left(Q_Y(F_\Delta(\delta|x)|x)|x\right)}\right]^{-1}. \tag{3.22}$$

Theorem 3 (Limit Distribution for estimators of the marginal distribution and quantile functions)**.** *Let* $\hat{F}_\Delta(\delta) = \int_{\mathcal{X}} \hat{F}_\Delta(\delta|x)dF_X^o(x)$ *be an estimator of the marginal distribution of the effects* $F_\Delta(\delta)$*. Under the conditions* ***M.1****,* ***M.2****,* ***C.1****,* ***Q.1****, and* ***R.P****, we have:*

$$\sqrt{n}\left(\hat{F}_\Delta(\delta) - F_\Delta(\delta)\right) \Rightarrow \int_{\mathcal{X}} Z_g(\delta, x)dF_X^o(x) := Z_g(\delta), \tag{3.23}$$

in the space $\ell^\infty(\mathcal{D})$*, where the Gaussian random function* $\delta \mapsto Z_g(\delta)$ *has zero mean and covariance function*

$$\Omega_Z(\delta, \tilde{\delta}) := \int_{\mathcal{X}}\int_{\mathcal{X}} \Omega_Z(\delta, x, \tilde{\delta}, \tilde{x})dF_X^o(x)dF_X^o(\tilde{x}). \tag{3.24}$$

Let $\hat{Q}_\Delta(u) = \inf\{\delta : \hat{F}_\Delta(\delta) \geq u\}$ *be an estimator of the conditional quantile function of the effects* $Q_\Delta(u)$*. Under the conditions* ***M.1****,* ***M.2****,* ***C.1****,* ***Q.1****, and* ***R.P****, we have:*

$$\sqrt{n}\left(\hat{Q}_\Delta(u) - Q_\Delta(u)\right) \Rightarrow -f_\Delta(Q_\Delta(u))^{-1}Z_g(Q_\Delta(u)) := V_g(u), \tag{3.25}$$

---

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

in the space $\ell^\infty((0,1))$, *where* $f_\Delta(\delta) = \int_{\mathcal{X}} f_\Delta(\delta|x) dF_X^o(x)$ *and the Gaussian random function* $u \mapsto V_g(u)$ *has zero mean and covariance function*

$$\Omega_V(u,\tilde{u}) := f_\Delta(Q_\Delta(u)) f_\Delta(Q_\Delta(\tilde{u})) \Omega_Z(Q_\Delta(u), Q_\Delta(\tilde{u})). \tag{3.26}$$

Example 2. Quantile regression. To illustrate the previous analysis, it is convenient to consider the linear quantile regression model where $Q_Y(u|x) = x'\beta(u)$. In this case, under suitable regularity conditions and *i.i.d.* sampling, the Koenker and Bassett (1978) quantile regression estimator satisfies

$$\sqrt{n}\left(\hat{\beta}(u) - \beta(u)\right) \Rightarrow J(u)^{-1} B(u), \tag{3.27}$$

where $B(u)$ is a zero mean Gaussian process with covariance function $(\min(u,\tilde{u}) - u \cdot \tilde{u}) E[XX']$ proportional to the covariance function of a Brownian bridge, and $J(u) = E[f_Y(Q_Y(u|X)|X)XX']$. Hence,

$$\sqrt{n}\left(\hat{Q}_Y(u|x) - Q_Y(u|x)\right) = \sqrt{n}\left(x'\hat{\beta}(u) - x'\beta(u)\right) \Rightarrow V(u,x) = x'J(u)^{-1}B(u), \tag{3.28}$$

with covariance function given by:

$$\Sigma_V(u,x,\tilde{u},\tilde{x}) = (\min(u,\tilde{u}) - u \cdot \tilde{u}) x' J(u)^{-1} E[XX'] J(\tilde{u})^{-1} \tilde{x}. \tag{3.29}$$

The covariance function for $\hat{Q}_Y^j(u)$ takes the form:

$$\Sigma_V^j(u,\tilde{u}) = \frac{\int\int f_Y(Q_Y^j(u)|x) f_Y(Q_Y^j(\tilde{u})|\tilde{x}) \Sigma_V(u,x,\tilde{u},\tilde{x}) dF_X^j(x) dF_X^j(\tilde{x})}{\left[\int f_Y(Q_Y^j(u)|x) dF_X^j(x)\right]^2}, \tag{3.30}$$

where

$$f_Y(y|x) = \left.\frac{1}{x'\partial\beta(u)/\partial u}\right|_{u=F_Y(y|x)} = \frac{1}{x'J(F_Y(y|x))^{-1}E[X]}, \tag{3.31}$$

see proof of Theorem 3 in Angrist, Chernozhukov, and Fernandez-Val (2006). Similar expressions can be obtained for the covariance functions of the estimators of other functionals of the marginal distributions of the outcomes and effects. In Appendix B we provide consistent estimators for the components of these expressions that can be used to perform pointwise inference about specific features of the policy effect in this model.

3.5. **Uncertainty about the distribution of the covariates.** The previous analysis assumes that the distributions of the covariates before and after the policy intervention were known in the population. In practice, however, we usually only observe a sample of the covariates and outcome before the intervention and a sample of the covariates after the intervention. In this case the previous limit distribution theory is still valid to make inference about the individuals in the sample, but in order to make inference about the entire population we need to take into account the additional source of variation coming from the estimation of the distributions of the covariates.

Let n/λ^j denote the sample size for the covariates before and after the policy intervention, where j indexes the status, $j \in \{o, c\}$ and we normalize $\lambda^o = 1$. We make the basic assumption that the estimator of the distribution function of the covariates $x \mapsto \hat{F}^j_X(x)$ converges in law to a Gaussian process:

$$\sqrt{n}\left(\hat{F}^j_X(x) - F^j_X(x)\right) \Rightarrow \lambda^j B^j_X(x), \tag{3.32}$$

in the space $\ell^\infty(\mathcal{X})$. The convergence holds jointly for all estimators indexed by the status $j \in \{o, c\}$. This assumption is not very restrictive as it is satisfied by the empirical distribution function under general sampling conditions.[5] The joint convergence holds trivially in the leading cases where the counterfactual distribution is a known transformation of the observed distribution, or when the two distributions are estimated from independent samples.

The estimation of the covariate distribution affects the inference processes for the functionals of interests. Take, for example, the marginal distribution functions. When the

[5]Under *i.i.d.* sampling, for example, B^j_X is a P_X-Brownian bridge with zero mean and covariance function $E[B^j_X(x)B^j_X(\tilde{x})] = F_X(\min(x, \tilde{x})) - F_X(x)F_X(\tilde{x})$, where the minimum is taken componentwise; see, e.g., Billingsley (1968) and Neuhaus (1971).

covariate distribution is unknown, a feasible estimator for these functions can be constructed as $\hat{F}_Y^j(y) = \int_{\mathcal{X}} \hat{F}_Y(y|x) d\hat{F}_X^j(x)$. The limit process for this estimator becomes:

$$\sqrt{n}\left(\hat{F}_Y^j(y) - F_Y^j(y)\right) \Rightarrow Z(y) + \sqrt{\lambda^j} \int_{\mathcal{X}} F_Y(y|x) dB_X^j(x), \tag{3.33}$$

where the first component comes from the estimation of the conditional model and the second comes from the estimation of the distribution of the covariates. These components are independent under correct specification of the conditional model leading to the following covariance function for the limit process under *i.i.d.* sampling:

$$\Sigma_Z(y, \tilde{y}) + \lambda^j \int_{\mathcal{X}} \left[F_Y(y|x) - F_Y^j(y)\right] \left[F_Y(\tilde{y}|x) - F_Y^j(\tilde{y})\right] dF_X^j(x). \tag{3.34}$$

Similar expressions can be obtained for the other functionals of interest. In Appendix C, we re-derive the main limit distribution results for the case where the covariate distributions are estimated.

3.6. **Uniform inference and resampling methods.** The previous limit distribution results can be readily used to perform inference on particular features of the distributions of the outcome before and after the policy. Thus, for example, a direct implication of Corollary 1 is that the quantile treatment effect estimator for a given quantile u is distributed asymptotically as normal with mean $QTE_Y(u)$ and variance $\Sigma_W(u,u)/n$. We can therefore routinely carry out pointwise inference on $QTE_Y(u)$ using the normal distribution by replacing the unknown components of $\Sigma_W(u,u)$ for consistent sample estimates.

Pointwise inference, however, only permits looking at specific aspects of the effect of the policy separately. This might be restrictive for policy analysis where the quantities and hypotheses of interest usually involve the entire distribution of the observed and counterfactual outcomes. Thus, for example, inference questions such as that the policy has no effect, has constant effect, or has beneficial effect cannot be tested by looking at specific quantiles of the outcome distribution. Moreover, simultaneous inference corrections to pointwise procedures based on the normal distribution, such as Bonferroni-type corrections, can be very conservative to perform multiple inference for highly dependent hypotheses. These procedures are also no longer suitable for testing a continuum of hypotheses. A convenient and computationally attractive alternative to perform inference on functions is to use Kolmogorov-type procedures based on the entire limit processes.

Kolmogorov-type inference is complicated in this case because the inference processes are non-pivotal, as their covariance functions depend on unknown, though estimable, nuisance parameters. Moreover, there does not seem to be a simple transformation to make these limit processes distribution-free. Similar non-pivotality issues arise in a variety of goodness-of-fit problems studied by Durbin and others, and are referred to by Koenker and Xiao (2002) as the Durbin problem. This problem makes analytical methods for inference difficult to implement. Thus, the limit distribution of the Kolmogorov test statistics can in principle be simulated replacing the covariance functions for uniformly consistent estimates. This simulation approach, however, can be cumbersome since the limit processes are non-standard and design-specific, so a new simulation is needed for each application.

A way to partially overcome the problems of the analytical approach is to use resampling methods. Our limit distribution results rely only on the compact differentiability of the functionals of interests with respect to the limit of the conditional processes, and on these conditional processes following a functional central limit theorem. An additional advantage of this general approach is that the compact differentiability preserves the validity of bootstrap to perform inference on the functional of interests, if the underlying conditional processes are bootstrap-able. This result, stated formally in the next corollary, follows directly from the functional delta method for the bootstrap, see, e.g., van der Vaart (1998).

Corollary 4 (Validity of bootstrap for uniform inference)**.** *If the conditional processes (3.1), (3.2), and (3.18) satisfy the conditions to guarantee the validity of bootstrap, then the limit processes (3.6), (3.9), (3.12), (3.14), (3.16), (3.23), and (3.25) also satisfy these conditions.*

The bootstrap procedure can be also computationally very intensive, but avoids estimating the components of the covariance functions. Moreover, if the sample size is large the computational complexity of the bootstrap procedure can be reduced by resampling the first order conditions of the estimators of the conditional models, see Parzen, Wei, and Ying (1994) and Chernozhukov and Hansen (2006); or by using subsampling, see Chernozhukov and Fernandez-Val (2006).

4. Illustrative Examples

4.1. **Empirical example.** To illustrate the applicability of the previous results to perform inference on counterfactual distributions we consider the estimation of expenditure curves. We use the Engel (1857) data set, originally collected by Ducpetiaux (1855) and Le Play (1855) from 235 budget surveys of 19th century working-class Belgium households, to estimate the relationship between food expenditure and annual household income (see also Perthel, 1975). Ernst Engel originally presented these data to support the hypothesis that food expenditure constitutes a declining share of household income (Engel's Law). Here, we estimate marginal quantile functions of food expenditure under different distributions for the annual household income.[6]

For our counterfactual exercise we consider two distributions of income: the observed distribution and a hypothetical redistribution. The redistribution consists of a neutral reallocation of income from above to below the mean that reduces the standard deviation of the observed income by 25%. This policy can be implemented by the progressive income tax

$$X^c = E[X^o] + .75(X^o - E[X^o]), \tag{4.1}$$

which yields a counterfactual distribution of income

$$F_X^c(x) = F_X^o(E[X^o] + (x - E[X^o])/.75), \tag{4.2}$$

[6]All the computations were carried out using the software R (R Development Core Team, 2007) and the quantile regression package quantreg (Koenker, 2007).

where F_X^o is the observed distribution of income in the data set. The observed and counterfactual distributions of income are plotted in Figure 1. Here we can see that the counterfactual distribution is a mean preserving spread of the observed distribution that reduces standard deviation by 25%, while keeping the mean constant.

We consider three different conditional models for the relationship between food expenditure and annual income: a linear location-shift model, a quantile regression model, and a distribution regression model. For the location model we estimate the location parameters by least squares and use the sample quantiles of the residuals to estimate the quantiles of the disturbance. For the linear quantile model, the entire conditional quantile function is estimated by running quantile regressions at several quantiles. For the distribution model, the conditional distribution is obtained by estimating logits of the indicator functions $1\{Y \leq y\}$ on the covariates for a grid of values of y corresponding to the values of food expenditure in the data set. The logit estimates are monotonized by rearrangement, what also produces estimates of the conditional quantile function.

To analyze the effect on the food expenditure distribution of our hypothetical "policy" exercise, Figures 2, 3, and 4 plot 90% simultaneous bands for the marginal quantile functions of food expenditure before and after the income redistribution based on the three different conditional models. These uniform bands, which allow us to perform inference on the functions without compromising the joint confidence level, are constructed using 500 bootstrap repetitions and a grid of quantiles $\{0.10, 0.11, ..., 0.90\}$. The left panels show estimates of the observed and counterfactual quantile functions of food expenditure. For the three conditional estimation methods, the redistribution has a slightly bigger effect on the lower tail of the expenditure distribution, but the confidence bands are only significantly different for a few quantiles in the quantile regression model. The right panel refines this finding by plotting 90% simultaneous bands for the quantile treatment effects. Here we can see more clearly that the policy has a positive impact on the lower tail of the expenditure distribution, and a negative effect on the upper tail. This result is consistent with a consumption pattern where food expenditure is highly correlated with income, and all the households adjust their levels of expenditure to the changes in income.

Since in this case we have a perfectly controlled experiment, we can also estimate the distribution of effects of the policy. Figure 5 plots 90% confidence bands for the quantiles of the effects based on the three conditional estimators. Here, we can see that the effects of the redistribution can be large and very heterogenous. Figure 6 shows that the progressive income redistribution reduces food expenditure inequality based on 90% confidence sets for Lorenz curves and the corresponding Gini coefficients.[7]

4.2. **Monte Carlo.** We conduct a Monte Carlo experiment, matching closely the previous empirical application, to illustrate the finite sample properties of the policy estimators. In particular, we consider a data generating process (DGP) based on a conditional location-scale shift model: $Y = Z(X)'\alpha + (Z(X)'\gamma)\epsilon$, where ϵ is independent of X, with true conditional quantile function

$$Q(u|X) = Z(X)'\alpha + (Z(X)'\gamma)Q_\epsilon(u).$$

The regressor vector $Z(X)$ includes a constant and a covariate X, namely $Z(X) = (1, X)'$. The observed distribution of X corresponds to the empirical distribution of income in the Engel data set; whereas the counterfactual distribution corresponds to a neutral income redistribution that reduces the standard deviation by 25% implemented as in (4.1). The parameters of the conditional model are set to $\alpha = (624.15, 0.55)$ and $\gamma = (1, 0.0013)$. These values are calibrated to match the Engel empirical example, employing the estimation method of Koenker and Xiao (2002).We draw 1,000 Monte Carlo samples of size $n = 235$ from the DGP. To generate the values of the observed outcome, we draw observations from a normal distribution with the same mean and variance as the residuals

[7]The Gini coefficient is twice the area between the 45^o line (line of perfect equality) and the Lorenz curve. It takes values between 0 and 1, with 0 corresponding to perfect equality and 1 corresponding to perfect inequality.

$\epsilon = (Y - Z(X)'\alpha)/(Z(X)'\gamma)$ of the Engel data set; and we draw values for the observed covariate X^o from the empirical distribution of income.

We consider three different estimators for the conditional model to assess the properties of the policy estimators under correct and incorrect specification. Thus, in each replication we estimate the conditional distribution or quantile function using a correctly specified linear quantile regression model and misspecified linear location-shift and logit regression models. The functionals of interest are the quantile functions of the observed and counterfactual outcome, the quantile treatment effects function, the quantile function of the distribution of the effects, Lorenz curves for the observed and counterfactual outcome distributions, and the corresponding Gini coefficients. For each of the conditional estimators considered, these functional are obtained using the procedure described in section 2.5, where the observed and counterfactual distributions of the covariates are estimated by the empirical distribution of income in each replication and by applying the transformation in (4.2) to this empirical distribution. We use bootstrap with 200 repetitions to obtain 90% confidence bands for the functionals of interest in each replication.

Table 1 reports measures of the bias and inference properties for the policy estimators. The integrated bias is obtained by Monte Carlo average of

$$\int \left| \hat{f}(t) - f_0(t) \right| dt, \tag{4.3}$$

where $f_0(t)$ is the true functional and $\hat{f}(t)$ is its policy estimator. Coverage probabilities correspond to Monte Carlo frequencies that the 90% bootstrap confidence bands include the entire functional of interest. The columns labeled as *Length 90% CI / 5-95 Range* give the integrated ratio of the Monte Carlo average length of the 90% confidence band to the Monte Carlo 5-95 quantile spread of the estimates.

Overall, the results in Table 1 show similar patterns to the empirical example with the location and quantile regression models giving more precise (relative to the 5-95 quantile spread) bands than the distribution regression for the quantile and quantile treatment effects functions. Misspecification of the conditional model introduces bias in the policy estimators, specially for the more restrictive location model. The coverage frequencies for the correctly specified quantile regression model are close to their nominal levels for the quantile and quantile treatment effects functions, even for a sample size of only 235 observations. The logit regression has also coverage frequencies close to the nominal levels as the misspecification bias is compensated by overestimation of the size of the confidence bands. The bands for the Lorenz curves and Gini coefficients have generally lower coverage than the nominal level.

5. Conclusion

This paper provides methods to make inference about the effect on an outcome of interest of a change in the distribution of policy-related variables. The validity of the proposed inference procedures in large sample relies only on the applicability of a functional central limit theorem for the estimator of the relationship between the outcome and covariates. This condition holds for common semiparametric estimators of conditional distribution and quantile functions, such as quantile regression and proportional hazard models. It would be interesting to extend the analysis to the case where the assumptions about the conditional model are relaxed by using nonparametric conditional distribution or quantile estimators. This extension is the object of current research by the authors.

Appendix A. Proofs

A.1. **Notation.** Define $Y_x := Q_Y(U|x)$, where $U \sim$ 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 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 $h : \mathcal{UX} \to \mathbb{R}$, $C(\mathcal{UX})$ denotes the set of continuous functions mapping $h : \mathcal{UX} \to \mathbb{R}$, and $\ell^1(\mathcal{UX})$ denotes the set of measurable functions $h : \mathcal{UX} \to \mathbb{R}$ such that $\int_{\mathcal{U}} \int_{\mathcal{X}} |h(u|x)| du dx < \infty$, where du and dx denote the integration with respect to the Lebesgue measure on $\mathcal{U}$ and $\mathcal{X}$, respectively.

A.2. **Auxiliary Lemmas.**

Lemma 3 (Equivalence between continuous convergence and uniform convergence)**.** *Let D and D' be complete separable metric spaces, with D compact. Suppose $f : D \to D'$ is continuous. Then a sequence of functions $f_n : D \to D'$ converges to f uniformly on D if and only if for any convergent sequence $x_n \to x$ in D we have that $f_n(x_n) \to f(x)$.*

Proof of Lemma 3: See, for example, Resnick (1987), page 2. □

Lemma 4 (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.1**, as $t \to 0$,*

$$D_{h_t}(y|x,t) = \frac{F_Y(y|x,h_t) - F_Y(y|x)}{t} \to D_h(y|x), \tag{A.1}$$

$$D_h(y|x) := -f_Y(y|x) h(F_Y(y|x)|x). \tag{A.2}$$

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 4: 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 $u \notin B_\epsilon(F_Y(y|x))$, as $t \to 0$,

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

Therefore,

$$\begin{aligned} & \frac{\int_0^1 1\{Q_Y(u|x) + th_t(u|x) \le y\}du - \int_0^1 1\{Q_Y(u|x) \le y\}du}{t} \\ \le & \int_{B_\epsilon(F_Y(y|x))} \frac{1\{Q_Y(u|x) + t(h(F_Y(y|x)|x) - \delta) \le y\} - 1\{Q_Y(u|x) \le y\}}{t} du, \end{aligned} \tag{A.3}$$

which by the change of variable $y' = Q_Y(u|x)$ is equal to

$$\frac{1}{t} \int_{J \cap [y, y - t(h(F_Y(y|x)|x) - \delta)]} f_Y(y'|x) dy',$$

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 $t \to 0$, 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 $f_Y(y'|x) \to f_Y(y|x)$ as $F_Y(y'|x) \to F_Y(y|x)$. Therefore, the right hand term in (A.3) 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 (A.3) 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 3. 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 $h(u|x)$ in both arguments, continuity of $F_Y(y|x)$, and the assumed uniform continuity of $f_Y(y|x)$ in both arguments. $\square$

A.3. **Proof of Lemma 1.** This Lemma simply follows by the functional delta method (e.g., van der Vaart, 1998) by the Hadamard differentiability of $F_Y(y|x)$ with respect to $Q_Y(u|x)$ shown in Lemma 4. Instead of restating what this method is, it takes less space to simply adapt the proof to the current context.

Consider the map $g_n(y,x|h) = \sqrt{n}(F_Y(y|x, h/\sqrt{n}) - F_Y(y|x))$. The sequence of maps satisfies $g_{n'}(y,x|h_{n'}) \to D_h(y|x)$ in $\ell^\infty(K)$ for every subsequence $h_{n'} \to h$ in $\ell^\infty(\mathcal{UX})$, where h is continuous. It follows by the Extended Continuous Mapping Theorem that, in $\ell^\infty(K)$, $g_n(y,x|\sqrt{n}(\hat{Q}(u|x) - Q_Y(u|x))) \Rightarrow D_V(y|x)$ as a stochastic process indexed by (y,x), since $\sqrt{n}(\hat{Q}(u|x) - Q_Y(u|x)) \Rightarrow V(u,x)$ in $\ell^\infty(\mathcal{UX})$. □

A.4. **Proof of Theorem 1.** The joint uniform convergence result follows from Condition **D.1** by the Extended Continuous Mapping Theorem, since the integral is a continuous operator. Gaussianity of the limit process follows from linearity of the integral. The derivation of the mean and covariance functions of the limit processes is standard and therefore we omit it. □

A.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, since the quantile operator is Hadamard differentiable (see, e.g., Fernholz, 1983, and Lemma 4). The derivation of the mean and covariance functions of the limit processes is standard and therefore we omit it. □

A.6. **Proof of Corollary 1.** This result follows directly from Theorem 2 by the Extended Continuous Mapping Theorem. The derivation of the mean and covariance function of the limit process is standard and therefore we omit it. □

A.7. **Proof of Corollary 2.** This result follows directly from Theorem 1 by the Extended Continuous Mapping Theorem. The derivation of the mean and covariance function of the limit process is standard and therefore we omit it. □

A.8. **Proof of Corollary 3.** This result follows directly from Theorem 1 by the Functional Delta Method. The derivation of the mean and covariance function of the limit process is standard and therefore we omit it. □

A.9. **Proof of Lemma 2.** The uniform convergence result for the conditional quantile process $\sqrt{n}\left(\hat{Q}_\Delta(u|x) - Q_\Delta(u|x)\right)$ follows from Conditions **Q.1** and **R.P.** by the Extended Continuous Mapping Theorem. Uniform convergence of the conditional distribution process $\sqrt{n}(\hat{F}_\Delta(\delta|x) - F_\Delta(\delta|x))$ follows from the covergence of the quantile process by Functional Delta Method. The Hadamard differentiability of $F_\Delta(\delta|x)$ with respect to $Q_\Delta(u|x)$ can be established using the same argument as in the proof of Lemma 4. The expression for $f_\Delta(\delta|x)$ follows from $Q_\Delta(u|x) = Q_Y(u|g(x)) - Q_Y(u|x)$, $Q'_Y(u|x) = 1/f_Y(Q_Y(u|x)|x)$, and the Inverse Function Theorem. The derivation of the mean and covariance functions of the limit processes is standard and therefore we omit it. □

A.10. **Proof of Theorem 3.** The uniform convergence result for the distribution function follows from the convergence of the conditional process in Lemma 2 by the Extended Continuous Mapping Theorem, 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, since the quantile operator is Hadamard differentiable (see, e.g., Fernholz, 1983, and Lemma 4). The derivation of the mean and covariance functions of the limit processes is standard and therefore we omit it. □

A.11. **Proof of Corollary 4.** This result follows from the Functional Delta Method for the Bootstrap (see, e.g., van der Vaart, 1998). □

Appendix B. Linear quantile regression model: pointwise inference

In order to make pointwise inference on the marginal quantile functions the components of the expression (3.30) need to be estimated. The difficulty here is to find estimators for

$F_Y(Q_Y^j(u)|x)$, $J(u)$, and $f_Y(Q_Y^j(u)|x)$ that are uniformly consistent in u. Chernozhukov, Fernandez-Val, and Galichon (2006) establishes the uniform consistency for the estimator of $F_Y(y|x)$

$$\hat{F}_Y(y|x) = \int_0^1 1\{x'\hat{\beta}(u) \leq y\} d\lambda(u), \tag{B.1}$$

where λ is a uniform measure over a fine enough grid over $(0,1)$. Then, $\hat{F}_Y^j(y) = \int_{\mathcal{X}} \hat{F}_Y(y|x) dF_X^j(x)$ is a uniformly consistent estimator of $F_Y^j(y)$ by the Extended Continuous Mapping Theorem, and $\hat{Q}_Y^j(u) = \inf\{y : \hat{F}_Y^j(y) \geq u\}$ is a uniformly consistent estimator for and $Q_Y^j(u)$ by the Functional Delta Method. For the Jacobian term, $J(u)$, Angrist, Chernozhukov, and Fernandez-Val (2006) establishes the uniform convergence of Powell's kernel estimator. Finally, the estimator

$$\hat{f}_Y(y|x) = \frac{1}{x'\hat{J}(\hat{F}_Y(y|x))^{-1}\bar{X}_n} \tag{B.2}$$

where $\bar{X}_n$ is the sample mean of the observed covariates and $\hat{J}(\cdot)$ is Powell's kernel estimator of the Jacobian, is uniformly consistent for the conditional density by the Extended Continuous Mapping Theorem.

Appendix C. Limit Distribution Theory: Estimated Covariate Distributions

We start by restating the Condition **D.1** to incorporate the assumptions about the estimators of the covariate distributions.

D.1' Let $\hat{Z}(y,x) := \sqrt{n}(\hat{F}_Y(y|x) - F_Y(y|x))$, and $\hat{B}_X^j(x) := \sqrt{n}(\hat{F}_X^j(x) - F_X^j(x))$ for $j \in \{o, c\}$. These processes converge jointly in law to a multivariate continuous Gaussian process having independent increments:

$$\left(\hat{Z}(y,x), \hat{B}_X^o(x), \hat{B}_X^c(x)\right) \Rightarrow \left(Z(y,x), B_X^o(x), \sqrt{\lambda^c} B_X^c(x)\right), \tag{C.1}$$

in the space $\ell^\infty(\mathcal{Y} \times \mathcal{X} \times \mathcal{X} \times \mathcal{X})$; where the random function $(y,x) \mapsto Z(y,x)$ has zero mean and uniformly bounded covariance function $\Sigma_Z(y,x,\tilde{y},\tilde{x}) := E[Z(y,x)Z(\tilde{y},\tilde{x})]$, the random functions $x \mapsto B_X^j(x)$, for $j \in \{o, c\}$, have zero means and uniformly bounded covariance functions $\Sigma_B^j(x, \tilde{x}) := E[B_X^j(x) B_X^j(\tilde{x})]$, $Z(y, x)$ is independent of $B_X^j(x)$, for $j \in \{o, c\}$, and $B_X^o(x)$ and $B_X^c(x)$ have uniformly bounded cross covariance function $\Sigma_B^{oc}(x, \tilde{x}) := E[B_X^o(x) B_X^c(\tilde{x})]$.

Theorem 4 (Limit Distribution for Marginal Distributions)**.** *Under conditions **M.1**, **M.2**, and **D.1'** the estimators of the marginal distribution functions converge in law to the following continuous Gaussian process:*

$$\sqrt{n}\left(\hat{F}_Y^j(y) - F_Y^j(y)\right) \Rightarrow Z^j(y) + \sqrt{\lambda^j} \int_{\mathcal{X}} F_Y(y|x) dB_X^j(x) := \tilde{Z}^j(y), \tag{C.2}$$

in the space $\ell^\infty(\mathbb{R})$, where the random function $y \mapsto \tilde{Z}^j(y)$ has zero mean and covariance function

$$\tilde{\Sigma}_Z^j(y, \tilde{y}) := \Sigma_Z^j(y, \tilde{y}) + \lambda^j E\left[\int_{\mathcal{X}} \int_{\mathcal{X}} F_Y(y|x) F_Y(\tilde{y}|\tilde{x}) dB_X^j(x) dB_X^j(\tilde{x})\right]. \tag{C.3}$$

The convergence holds jointly for all estimators indexed by the status $j \in \{o, c\}$, with cross covariance function

$$\tilde{\Sigma}_Z^{oc}(y, \tilde{y}) := \Sigma_Z^{oc}(y, \tilde{y}) + \sqrt{\lambda^c} E\left[\int_{\mathcal{X}} \int_{\mathcal{X}} F_Y(y|x) F_Y(\tilde{y}|\tilde{x}) dB_X^o(x) dB_X^c(\tilde{x})\right]. \tag{C.4}$$

Remark 2. *The expression for the second term of the covariance and cross covariance functions can be characterized for some leading cases:*

(1) *The distributions of the observed and counterfactual covariates correspond to different populations and are estimated by the empirical distributions using mutually independent random samples. In this case $B_X^j(x)$, for $j \in \{o, c\}$, are independent Brownian bridges. The second component of the covariance functions are*

$$\int_{\mathcal{X}} \left[F_Y(y|x) - F_Y^j(y)\right] \left[F_Y(\tilde{y}|x) - F_Y^j(\tilde{y})\right] dF_X^j(x), \tag{C.5}$$

for $j \in \{o, c\}$, and the second component of the cross covariance function is zero.

(2) *The counterfactual covariates are known transformations of the observed covariates, $X^c = g(X^o)$, and the observed covariate distribution is estimated by the*

empirical distribution from a random sample. In this case $B_X^o(x)$ and $B_X^j(x)$ are highly dependent Brownian bridges. The second components of the covariance functions are

$$\int_{\mathcal{X}} \left[F_Y(y|x) - F_Y^o(y)\right]\left[F_Y(\tilde{y}|x) - F_Y^o(\tilde{y})\right] dF_X^o(x), \tag{C.6}$$

for the observed outcome,

$$\int_{\mathcal{X}} \left[F_Y(y|g(x)) - F_Y^c(y)\right]\left[F_Y(\tilde{y}|g(x)) - F_Y^c(\tilde{y})\right] dF_X^o(x), \tag{C.7}$$

for the counterfactual outcome, and the second component of the cross covariance function is

$$\int_{\mathcal{X}} \left[F_Y(y|x) - F_Y^o(y)\right]\left[F_Y(\tilde{y}|g(x)) - F_Y^c(\tilde{y})\right] dF_X^o(x). \tag{C.8}$$

Proof of Theorem 4: The stochastic process for the distribution function $\sqrt{n}(\hat{F}_Y^j(y) - F_Y^j(y))$ can be decomposed in three components:

$$\int_{\mathcal{X}} \hat{Z}(y,x) dF_X^j(x) + \int_{\mathcal{X}} F_Y(y|x) d\hat{B}_X^j(x) - \frac{1}{\sqrt{n}} \int_{\mathcal{X}} \hat{Z}(y,x) d\hat{B}_X^j(x). \tag{C.9}$$

The first component converges uniformly to $Z(y)^j$ by Theorem 1. The uniform convergence of the second component to a Gaussian process follows from the convergence of empirical stochastic integrals by condition **D.1'** (see, e.g., DeJong and Davidson, 2000), and standard Ito's results for stochastic integrals of Gaussian processes since $F_Y(y|x)$ is uniformly continuous in both arguments. The third term is of order $o_p(1)$ uniformly in y since the stochastic integral $\int_{\mathcal{X}} \hat{Z}(y,x) d\hat{B}_X^j(x)$ is bounded in probability uniformly in y by condition **D.1'** (see, e.g., DeJong and Davidson, 2000). This result follows because the random function $Z(y,x)$ is square integrable uniformly in y (see, e.g., Proposition 7.41 in White (2001)), since $\int_{\mathcal{X}} \int_{\mathcal{X}} \Sigma_Z(x,y,\tilde{x},\tilde{y}) dx d\tilde{x}$ is bounded uniformly in $(y,\tilde{y})$ by condition **D.1'**. $\square$

Theorem 5 (Limit Distribution for Marginal Quantiles)**.** *Under the conditions* ***M.1****,* ***M.2****,* ***C.1****, and* ***D.1'*** *the estimators of the marginal quantile functions converge in law* *to the following continuous linear Gaussian functional:*

$$\sqrt{n}\left(\hat{Q}_Y^j(u) - Q_Y^j(u)\right) \Rightarrow -f_Y^j(Q_Y^j(u))^{-1}\tilde{Z}^j(Q_Y^j(u)) := \tilde{V}^j(u). \tag{C.10}$$

in the space $\ell^\infty((0,1))$*, where* $f_Y^j(y) = \int_{\mathcal{X}} f_Y(y|x)dF_X^j(x)$ *and the random function* $u \mapsto \tilde{V}^j(u)$ *has zero mean and covariance function*

$$\tilde{\Sigma}_V^j(u,\tilde{u}) := f_Y^j(Q_Y^j(u))^{-1} f_Y^j(Q_Y^j(\tilde{u}))^{-1}\tilde{\Sigma}_Z^j(Q_Y^j(u), Q_Y^j(\tilde{u})). \tag{C.11}$$

The convergence holds jointly for all estimators indexed by the status $j \in \{o,c\}$*, with cross-covariance function*

$$\tilde{\Sigma}_V^{oc}(u,\tilde{u}) := f_Y^o(Q_Y^o(u))^{-1} f_Y^c(Q_Y^c(\tilde{u}))^{-1}\tilde{\Sigma}_Z^{oc}(Q_Y^o(u), Q_Y^c(\tilde{u})). \tag{C.12}$$

Proof of Theorem 5: The joint uniform convergence result and Gaussianity of the limit process follow from Theorem 4 by the Functional Delta Method, since the quantile operator is Hadamard differentiable (see, e.g., Fernholz, 1983, and Lemma 4). The derivation of the mean and covariance functions of the limit processes is standard and therefore we omit it. □

Corollary 5 (Limit Distribution for Quantile treatment Effects)**.** *Under the conditions* ***M.1****,* ***M.2****,* ***C.1****, and* ***D.1'*** *the estimators of the quantile treatment effects converge in law to the following linear functional of continuous Gaussian processes:*

$$\sqrt{n}\left(\widehat{QTE}_Y(u) - QTE_Y(u)\right) \Rightarrow \tilde{V}^c(u) - \tilde{V}^o(u) := \tilde{W}(u), \tag{C.13}$$

in the space $\ell^\infty((0,1))$*, where the random function* $u \mapsto \tilde{W}(u)$ *has zero mean and covariance function*

$$\tilde{\Sigma}_W(u,\tilde{u}) := \tilde{\Sigma}_V^o(u,\tilde{u}) + \tilde{\Sigma}_V^c(u,\tilde{u}) - \tilde{\Sigma}_V^{oc}(u,\tilde{u}) - \tilde{\Sigma}_V^{oc}(\tilde{u},u). \tag{C.14}$$

Proof of Corollary 5: This result follows directly from Theorem 5 by the Extended Continuous Mapping Theorem. The derivation of the mean and covariance function of the limit process is standard and therefore we omit it. □

Corollary 6 (Limit Distribution for Distribution Effects)**.** *Under the conditions **M.1**, **M.2**, and **D.1** the estimator of the distribution effects converges in law to the following linear functional of continuous Gaussian processes:*

$$\sqrt{n}\left(\widehat{DE}_Y(y) - DE_Y(y)\right) \Rightarrow \tilde{Z}^c(u) - \tilde{Z}^o(u) := \tilde{S}(y), \tag{C.15}$$

in the space $\ell^\infty(\mathcal{Y})$, where the random function $y \mapsto \tilde{S}(y)$ has zero mean and covariance function

$$\tilde{\Sigma}_S(y, \tilde{y}) := \tilde{\Sigma}_Z^o(y, \tilde{y}) + \tilde{\Sigma}_Z^c(y, \tilde{y}) - \tilde{\Sigma}_Z^{oc}(y, \tilde{y}) - \tilde{\Sigma}_Z^{oc}(\tilde{y}, y). \tag{C.16}$$

Proof of Corollary 6: This result follows directly from Theorem 4 by the Extended Continuous Mapping Theorem. The derivation of the mean and covariance function of the limit process is standard and therefore we omit it. □

Corollary 7 (Limit Distribution for Differentiable Functionals)**.** *Let $H_Y(y) = \phi(F_Y^o, F_Y^c, y)$ be a Hadamard differentiable functional in the first two arguments, with derivatives ϕ'_o and ϕ'_c with respect to the first and second argument. Under the conditions **M.1**, **M.2**, and **D.1'** the estimator of the functional $H_Y(y)$ defined in (2.17) converges in law to the following linear functional of continuous Gaussian processes:*

$$\sqrt{n}\left(\hat{H}_Y(y) - H_Y(y)\right) \Rightarrow \phi'_o(F_Y^o, F_Y^c, y)\tilde{Z}^o(y) + \phi'_c(F_Y^o, F_Y^c, y)\tilde{Z}^c(y) := \tilde{T}(y), \tag{C.17}$$

in the space $\ell^\infty(\mathcal{Y})$, where the random function $y \mapsto \tilde{T}(y)$ has zero mean and covariance function

$$\tilde{\Sigma}_T(y, \tilde{y}) := \phi'_o\tilde{\phi}'_o\tilde{\Sigma}_Z^o(y, \tilde{y}) + \phi'_c\tilde{\phi}'_c\tilde{\Sigma}_Z^c(y, \tilde{y}) + \phi'_o\tilde{\phi}'_c\tilde{\Sigma}_Z^{oc}(y, \tilde{y}) + \tilde{\phi}'_o\phi'_c\tilde{\Sigma}_Z^{oc}(\tilde{y}, y), \tag{C.18}$$

where $\phi'_j := \phi'_j(F_Y^o, F_Y^c, y)$ and $\tilde{\phi}'_j := \phi'_j(F_Y^o, F_Y^c, \tilde{y})$, for $j \in \{o, c\}$.

Proof of Corollary 7: This result follows from the Functional Delta Method (see, e.g., van der Vaart, 1998). □

Corollary 8 (Validity of bootstrap for uniform inference)**.** *If the limit process in (C.1) satisfies the conditions to guarantee the validity of bootstrap, then the limit processes (C.2), (C.10), (C.13), (C.15), and (C.17) also satisfy these conditions.*

Proof of Corollary 8: This result follows from the Functional Delta Method for the Bootstrap (see, e.g., van der Vaart, 1998). □

Appendix D. Verification of regularity conditions for common conditional model estimators

Example 1. Location regression. Consider the linear location regression model $Y = X'\beta + V$, where the disturbance V is independent of X, with mean zero, variance σ_V^2 and quantile function $Q_V(u)$. In this case, the location parameter β can be estimated by OLS and the quantiles of V can be estimated by the sample quantiles of the OLS residuals. The estimator of the conditional cdf of Y is therefore

$$\hat{F}_Y(y|x) = \hat{F}_{\hat{V}}(y - x'\hat{\beta}). \tag{D.1}$$

Under suitable regularity conditions and *i.i.d.* sampling, by Theorem 2 in Durbin (1973) we have

$$\sqrt{n}\left(\hat{F}_{\hat{V}}(y - x'\hat{\beta}) - F_V(y - x'\beta)\right) \Rightarrow Z(y, x), \tag{D.2}$$

in the space $\ell^\infty(\mathbb{R} \times \mathcal{X})$, where $Z(y, x)$ is a Gaussian process with covariance function

$$\Sigma_Z(y, x, \tilde{y}, \tilde{x}) = [\min(F_Y(y|x), F_Y(\tilde{y}|\tilde{x})) - F_Y(y|x)F_Y(\tilde{y}|\tilde{x})] - f_Y(y|x)f_Y(\tilde{y}|\tilde{x})\sigma_V^2 x' E[XX']^{-1}\tilde{x}, \tag{D.3}$$

since $F_Y(y|x) = F_V(y - x'\beta)$ and $f_Y(y|x) = f_V(y - x'\beta)$.

Example 4. Distributional Regression. Consider the conditional model for the distribution function $F_Y(y|x) = \Lambda(x'\beta(y))$, where Λ is a known link function (logit or probit). The function $\beta(y)$ can be estimated by running logit or probit regressions of indicator variables $1\{Y \leq y\}$ on the regressor vector X (see, e.g., Foresi and Peracchi, 1995).

Under *i.i.d.* sampling and other regularity conditions, the estimator of $\beta(y)$ satisfies, in the space $\ell^\infty(\mathbb{R})$,

$$\sqrt{n}\left(\hat{\beta}(y) - \beta(y)\right) \Rightarrow -J(y)^{-1}B(y), \tag{D.4}$$

where $J(y) = E[\Lambda'(X'\beta(y))^2 XX'/\{\Lambda(X'\beta(y))(1-\Lambda(X'\beta(y)))\}]$, $\Lambda'(\cdot)$ is the derivative of $\Lambda(\cdot)$, and $B(y)$ is a zero mean Gaussian process with covariance function

$$\Sigma_B(y,\tilde{y}) = E\left[\frac{\Lambda'(X'\beta(y))\Lambda'(X'\beta(\tilde{y}))}{\Lambda(X'\beta(y))(1-\Lambda(X'\beta(\tilde{y})))}XX'\right], \tag{D.5}$$

for $\tilde{y} > y$. Hence,

$$\sqrt{n}\left(\hat{F}_Y(y|x) - F_Y(y|x)\right) \Rightarrow Z(y,x) = -\Lambda'(x'\beta(y))x'J(y)^{-1}B(y), \tag{D.6}$$

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

$$\Sigma_Z(y,x,\tilde{y},\tilde{x}) = \Lambda'(x'\beta(y))\Lambda'(\tilde{x}'\beta(\tilde{y}))x'J(y)^{-1}\Sigma_B(y,\tilde{y})J(\tilde{y})^{-1}\tilde{x}. \tag{D.7}$$

For example, in the case of the logit $\Lambda(u)' = \Lambda(u)(1-\Lambda(u))$. The asymptotic variance of the estimator of the marginal distribution based on the conditional model is

$$\begin{aligned}\Sigma_Z(y,y) &= E[\Lambda'(X'\beta(y))X]'E[\Lambda'(X'\beta(y))XX']^{-1}E[\Lambda'(X'\beta(y))X] \\ &= E[\Lambda(X'\beta(y))(1-\Lambda(X'\beta(y)))]. \end{aligned} \tag{D.8}$$

If the covariate distribution is estimated then the asymptotic variance becomes

$$E[\Lambda(X'\beta(y))(1-\Lambda(X'\beta(y)))] + E\left[(\Lambda(X'\beta(y)) - E[\Lambda(X'\beta(y))])^2\right], \tag{D.9}$$

which is the same as the asymptotic variance of the empirical marginal distribution function of Y. The logit estimates indeed not only have the same asymptotic variance but are also numerically identical to the empirical distribution estimates since

$$\hat{F}_Y(y) = \frac{1}{n}\sum_{i=1}^{n}\Lambda(X_i'\hat{\beta}(y)) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{1}\{Y_i \le y\}, \tag{D.10}$$

where the last equality follows from the first order conditions of the logit if the regressor X includes a constant term. Note that this result holds regardless of whether the conditional model is correctly specified or not.

References

[1] Albrecht, J., A. van Vuuren, and S. Vroman (2006): "Counterfactual Distributions with Sample Selection Adjustments: Econometric Theory and an Application to the Netherlands," preprint.

[2] Andrews, D.W.K (1988): "Chi-Square Diagnostic Tests for Econometric Models: Theory," *Econometrica* 56, pp. 1419-1453.

[3] Andrews, D.W.K. (1993): "Tests for Parameter Instability and Structural Change with Unknown Change Point," *Econometrica* 61, pp. 821-856.

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

[6] Billingsley, P. (1968): *Convergence of Probability Measures.* Wiley. New York.

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

[8] Chernozhukov, V. (2005): "Inference on Counterfactual Distributions Using Conditional Quantile Models," mimeo, MIT.

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

[10] Chernozhukov, V., I. Fernandez-Val and A. Galichon (2006): "Quantile and Probability Curves without Crossing," mimeo, MIT.

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

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

[13] DeJong, R.M., and J. Davidson (2000): *The functional Central Limit Theorem And Weak Convergence To Stochastic Integrals I: Weakly Dependent Processes,* Econometric Theory 16, pp. 621-642.

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

[15] 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.

[16] Ducpetiaux, E. (1855), *Budgets Économiques des Classes Ouvrières en Belgique.* Bruxelles.

[17] Engel, E. (1857), "Die Produktions und Konsumptionsverhältnisse des Königreichs Sachsen," *Zeitschrift des Statistischen Bureaus des Königlich Sächsischen Misisteriums des Innerm,* 8, pp. 1-54.

[18] Fernholz, L.T. (1983): *von Mises Calculus for Statistical Functionals,* Lecture Notes in Statistics 19, Springer, New York.

[19] Firpo, S., N. Fortin, and T. Lemieux (2007): "Unconditional Quantile Regressions," NBER Working Paper T0339.

[20] 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.

[21] 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.

[22] 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.

[23] 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.

[24] 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(4), pp. 487-535.

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

[26] Koenker, R. (2007). *quantreg: Quantile Regression.* R package version 4.10. `http://www.r-project.org`.

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

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

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

[30] Le Play, F. (1855), *Les Ouvriers Européens. Etudes sur les traveaux, la vie domestique et la condition morale des populations ouvrières de l'Europe.* Paris.

[31] Lehmann, E. (1974): *Nonparametrics: Statistical Methods Based on Ranks*, San Francisco: Holden-Day.

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

[33] Melly, B. (2006): "Estimation of Counterfactual Distributions Using Quantile Regression," preprint.

[34] Neuhaus, G. (1971): "On the Weak Convergence of Stochastic Processes with Multidimensional Time Parameter," *The Annals of Mathematical Statistics* 42(4), pp. 1285-1295.

[35] Owen, A.B. (1987): "Nonparametric Conditional Estimation," Technical Report 25, Laboratory for Computational Statistics, Stanford University, Dept. of Statistics.

[36] Parzen, M.I., L.J. Wei, and Z. Ying (1994): "A Resampling Method Based on Pivotal Estimating Functions," *Biometrika* 81, pp. 341-350.

[37] Perthel, D. (1975), "Engel's Law Revisited," *International Statistical Review* 43 , no. 2, 211–218.

[38] R Development Core Team (2007): *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL `http://www.R-project.org`.

[39] 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.

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

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

[42] Stocks, 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.

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

[44] Vaart, A. van der, and J. Wellner (1996), *Weak convergence and empirical processes: with applications to statistics*, New York: Springer.

[45] White, H. (2001): *Asymptotic Theory for Econometricians: Revised Edition*, Academic Press.

Table 1: Bias and Coverage Probabilities for Policy Estimators

	Integrated bias			Coverage (90% CI)			Length 90% CI / 5-95 Range		
	Location	QR	Logit	Location	QR	Logit	Location	QR	Logit
$Q_Y^o(u)$	20.24	15.30	17.40	60	87	90	1.32	1.37	1.83
$Q_Y^c(u)$	18.08	14.44	17.14	59	89	88	1.27	1.29	1.73
$QTE_Y(u)$	4.46	3.74	7.72	80	92	99	1.60	1.70	2.76
$Q_\Delta(u)$	4.28	3.51	4.48	93	89	74	1.96	1.71	1.76
$L_Y^o(y)$	0.012	0.008	0.010	11	74	86	1.34	1.44	5.76
$L_Y^c(y)$	0.010	0.006	0.007	9	73	80	1.34	1.44	4.41
Gi_Y^o	0.016	0.014	0.019	85	75	58	0.85	0.93	0.95
Gi_Y^c	0.013	0.011	0.013	85	74	67	0.83	0.92	0.98

Notes: 1000 Monte Carlo replications. 200 bootstrap repetitions in each replication. The design is a location-scale shift model calibrated to the Engel dataset.

4/13/2008

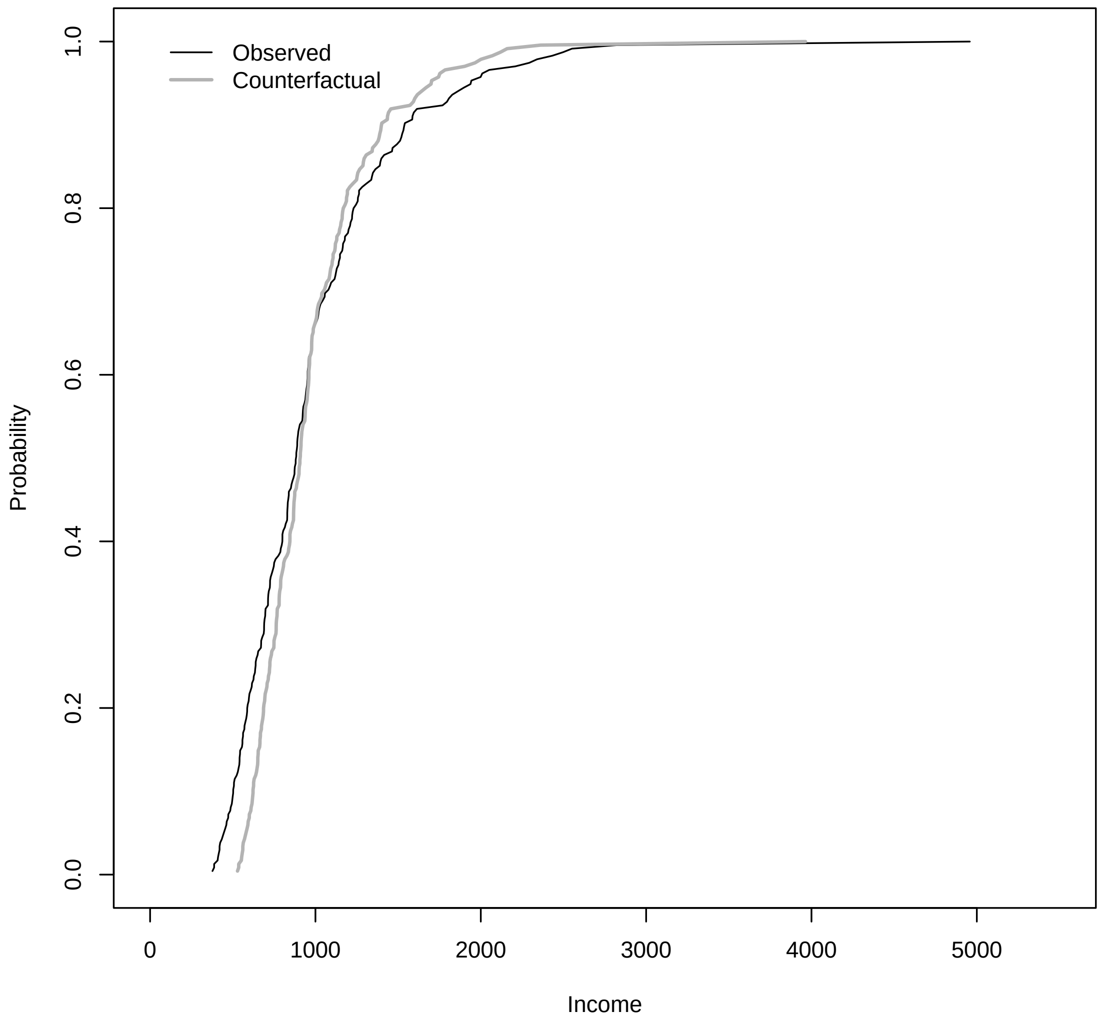

FIGURE 1. Observed and counterfactual empirical distribution for household income for the Engel food expenditure data. The counterfactual distribution is constructed as a neutral reallocation of the observed income from above to below the mean such that yields a 25% reduction in the standard deviation, that is $F_X^c(x) = F_X(\mu_X + (x - \mu_X)/.75)$.

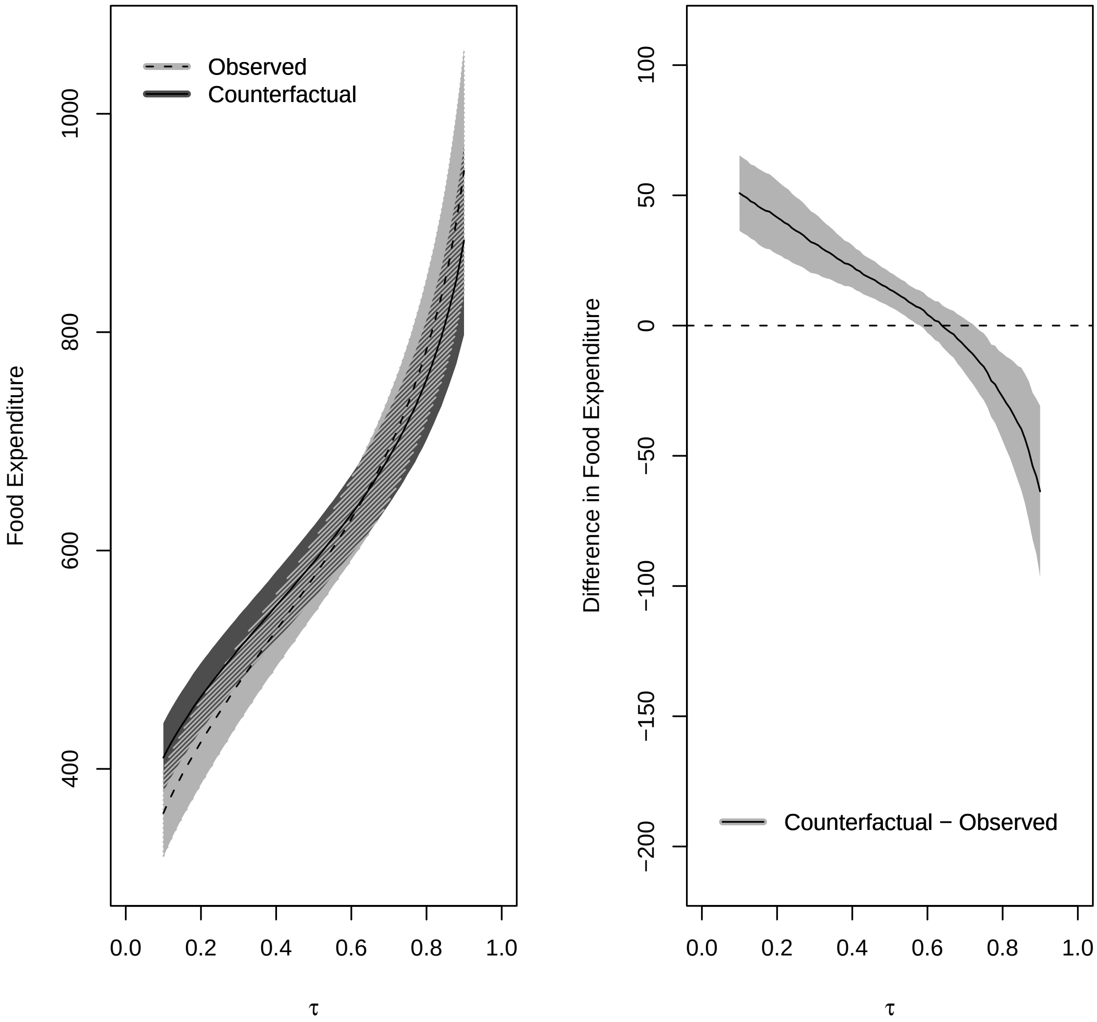

FIGURE 2. Simultaneous 90% confidence bands for quantile functions using the Engel food expenditure data: Location-shift model. Left panel shows uniform bands for the observed and counterfactual quantile functions of food expenditure. Counterfactual exercise consists of a mean-preserving spread of income that reduces standard deviation by 25%. Right panel plots uniform bands for quantile treatment effects.

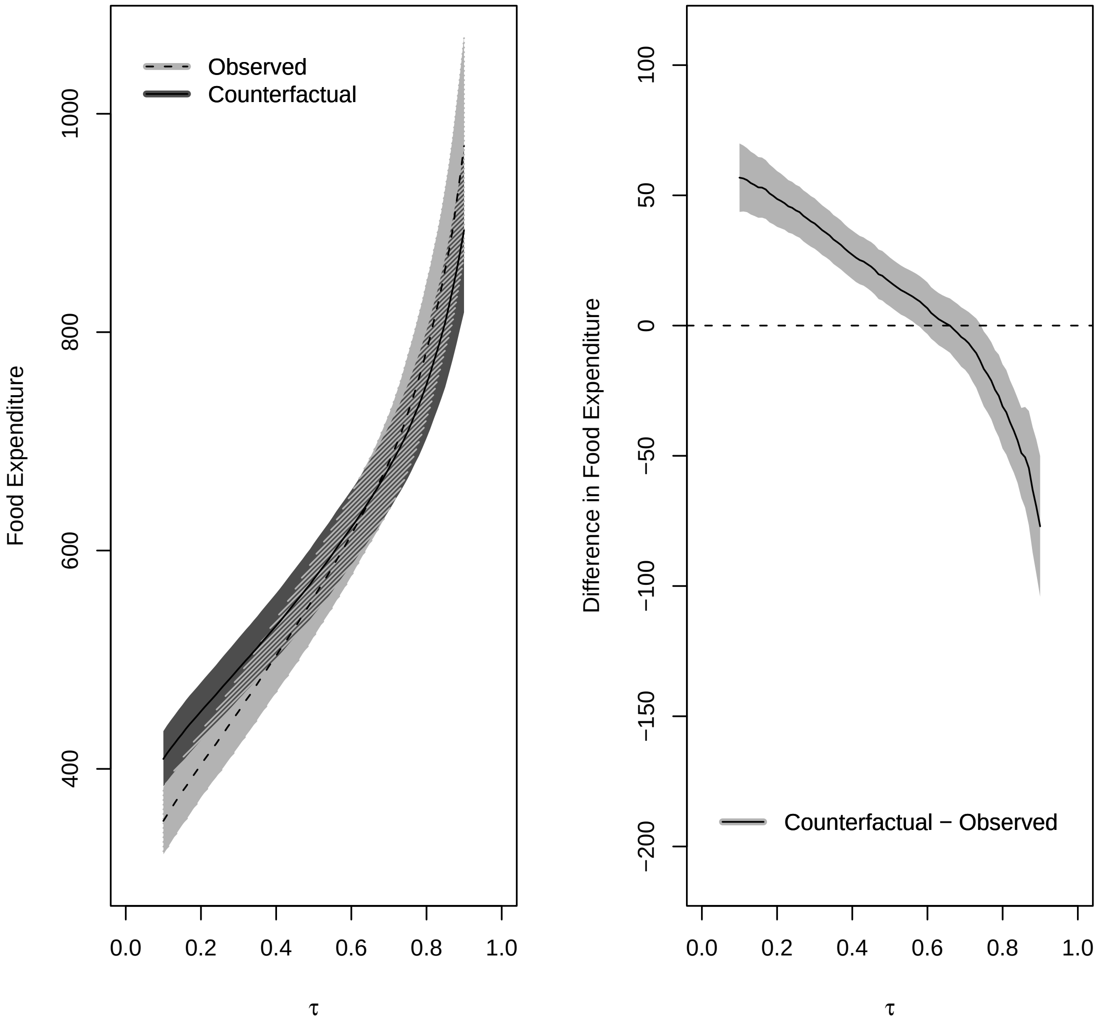

FIGURE 3. Simultaneous 90% confidence bands for quantile functions using the Engel food expenditure data: Quantile regression model. Left panel shows uniform bands for the observed and counterfactual quantile functions of food expenditure. Counterfactual exercise consists of a mean-preserving spread of income that reduces standard deviation by 25%. Right panel plots uniform bands for quantile treatment effects.

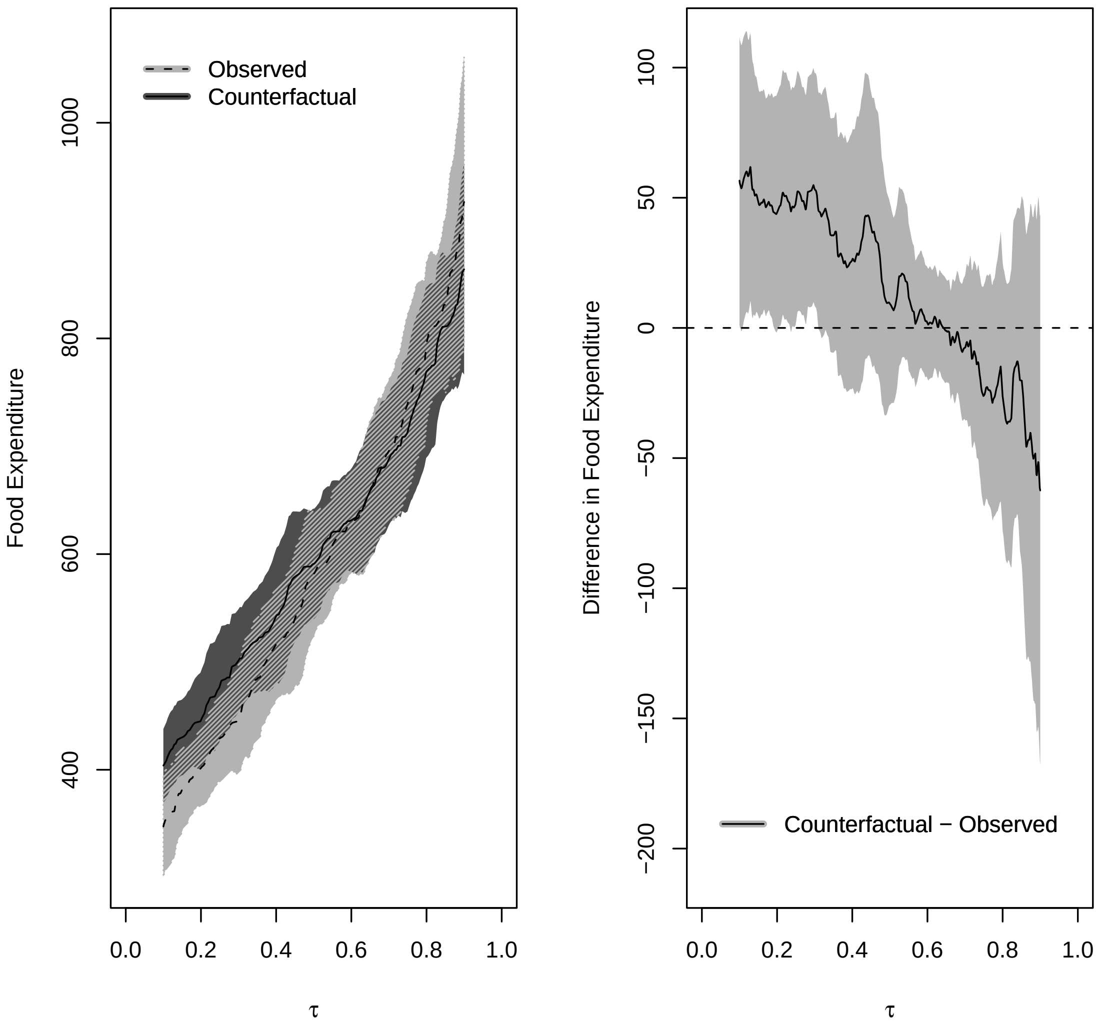

FIGURE 4. Simultaneous 90% confidence bands for quantile functions using the Engel food expenditure data: Distribution regression model. Left panel shows uniform bands for the observed and counterfactual quantile functions of food expenditure. Counterfactual exercise consists of a mean-preserving spread of income that reduces standard deviation by 25%. Right panel plots uniform bands for quantile treatment effects.

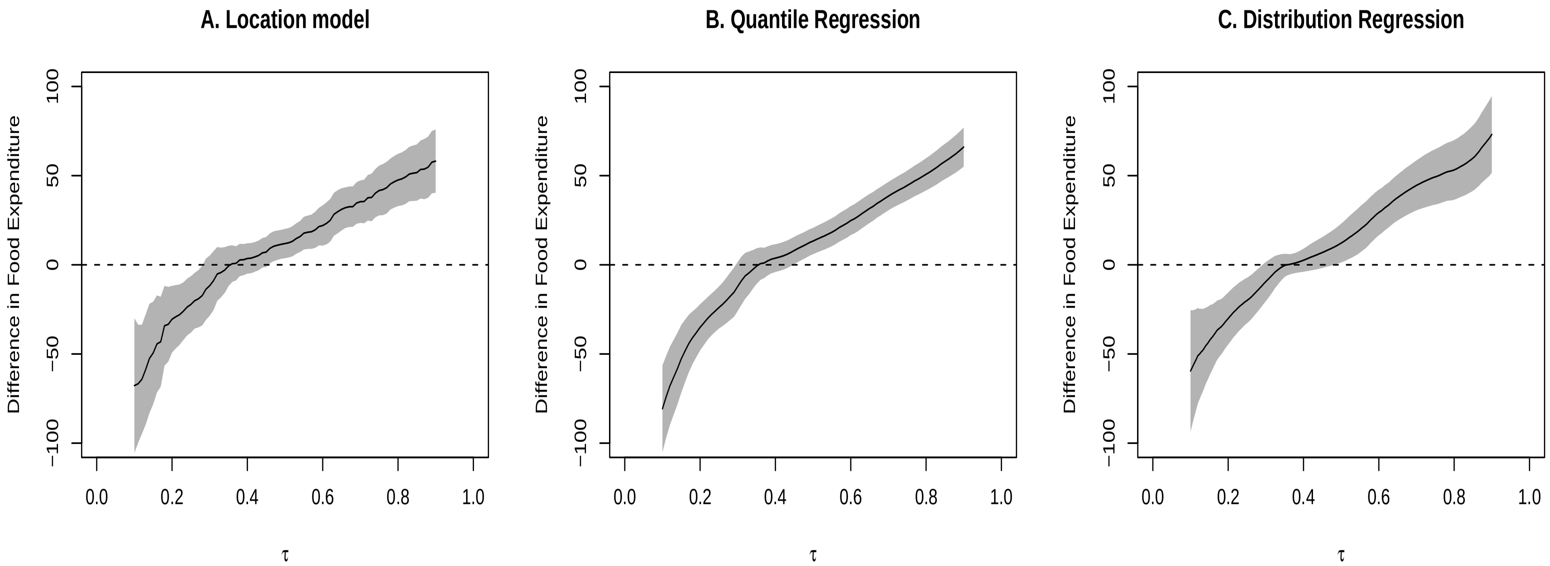

FIGURE 5. Simultaneous 90% confidence bands for the effect of the policy using the Engel food expenditure data. Counterfactual exercise consists of a mean-preserving spread of income that reduces standard deviation by 25%.

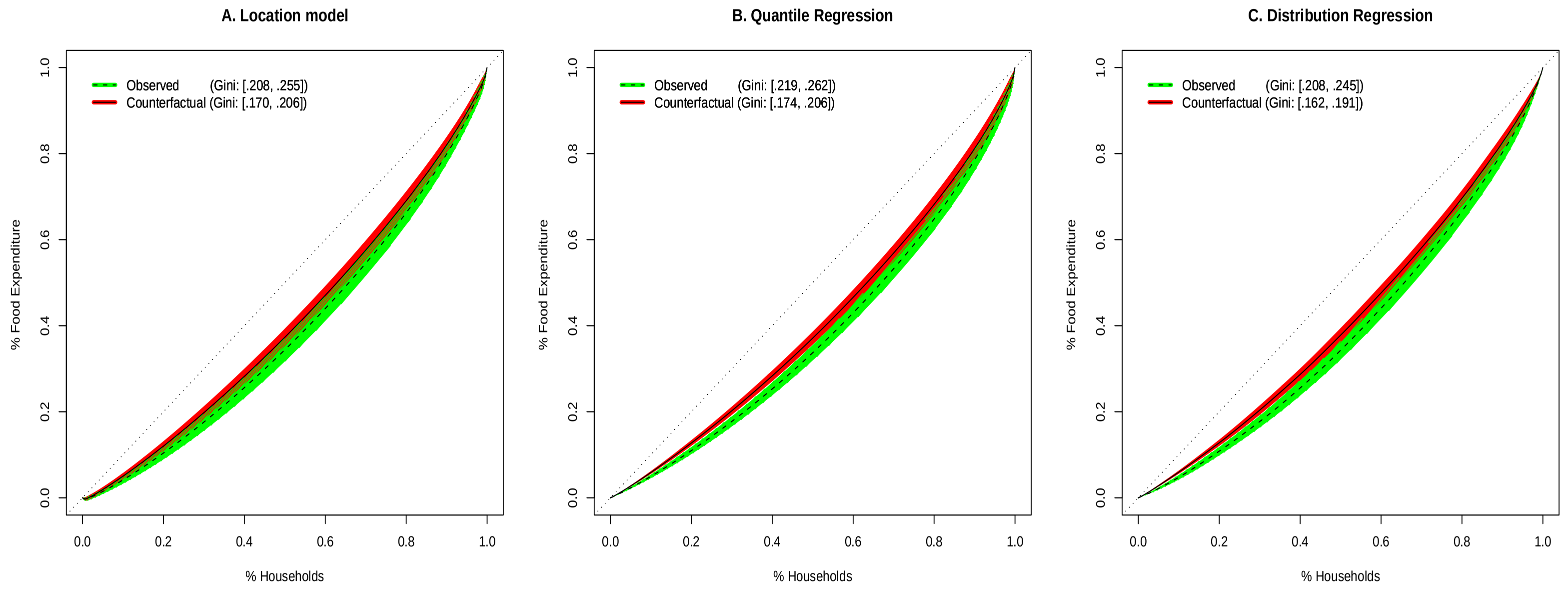

FIGURE 6. Simultaneous 90% confidence bands for Lorenz curves and 90% confidence intervals for Gini Coefficients using the Engel food expenditure data. Counterfactual exercise consists of a mean-preserving spread of income that reduces standard deviation by 25%.