

# NONPARAMETRIC ESTIMATION OF DISTRIBUTIONAL POLICY EFFECTS

Christoph Rothe\*

Toulouse School of Economics

First Version: April 9, 2008

This Version: September 15, 2009

## Abstract

This paper proposes a fully nonparametric procedure to evaluate the effect of a counterfactual change in the distribution of some covariates on the unconditional distribution of an outcome variable of interest. In contrast to other methods, we do not restrict attention to the effect on the mean. In particular, our method can be used to conduct inference on the change of the distribution function as a whole, its moments and quantiles, inequality measures such as the Lorenz curve or Gini coefficient, and to test for stochastic dominance. The practical applicability of our procedure is illustrated via a simulation study and an empirical example.

**JEL Classification:** C14, C31

**Keywords:** *Policy Effects, Counterfactual distribution, Nonseparable Model, Empirical Process, Nonparametric Regression*

---

\* Address: Toulouse School of Economics, 21 allée de Brienne, 31000 Toulouse, France. E-Mail: rothe@cict.fr. This paper is part of my PhD thesis at the University of Mannheim. I would like to thank my advisor Enno Mammen for his encouragement and support. I received further helpful comments from Richard Blundell, Andrew Chesher, Markus Frölich, Thomas Lemieux, Melanie Schienle, Kyusang Yu, an Associate Editor, two anonymous referees and seminar participants at Mannheim, UCL, the ESNASM 2009 Boston, ESEWM 2008 Cambridge, EEA-ESEM 2008 Milan and SETA 2008 Seoul. The usual disclaimer applies.

# 1 Introduction

In this paper, we propose a fully nonparametric procedure to evaluate the effect of a change in the distribution of some covariates on the unconditional distribution of an outcome variable of interest. We consider a general nonseparable model of the form

$$Y = m(X, \varepsilon) \quad (1.1)$$

where  $Y$  is the dependent variable of interest,  $X$  a vector of regressors and  $\varepsilon$  an unobserved error term that will usually represent individual heterogeneity. The question we are interested in is: How would the *unconditional* distribution of the dependent variable change if a policy maker could exogenously shift the values of  $X$  to some  $X^*$ , i.e what is the difference between the distribution of  $Y$  and the one of the (counterfactual) random variable

$$Y^* = m(X^*, \varepsilon).$$

We will call any difference between the distribution of  $Y$  and  $Y^*$  a *distributional policy effect*.

There are numerous examples in applied economics that fit into this rather abstract framework: Ichimura and Taber (2002) study the effect of a change in distribution of income induced by the introduction of a tuition subsidy on college attendance rates and future earnings; Stock (1991) considers the effects of cleaning up a nearby hazardous waste disposal site on average house prices; DiNardo, Fortin, and Lemieux (1996) analyse how the distribution of wages would have evolved in the United States between 1973 and 1992 if the distribution of workers' characteristics had remained at their 1973 level (see also Donald, Green, and Paarsch (2000), Gosling, Machin, and Meghir (2000) or Machado and Mata (2005)); and Blau and Kahn (1997) consider how much of the gender wage gap would persist if women had the same observable characteristics as men.

The contribution of our paper is to provide a fully nonparametric method to analyse these kind of questions. In contrast to other methods, we neither impose any parametric restrictions on the relation between  $Y$  and  $X$ , nor do we restrict attention to the policy's effect on the mean. In particular, our results can be used to conduct inference on the the change of the distribution function as a whole, its moments and quantiles, inequality measures such as the Lorenz curve or Gini coefficient, and to test for stochastic dominance. We show that the cumulative distribution function (CDF) of the counterfactual random variable  $Y^*$  is identified under some weak restrictions, and propose a two-stage estimator, that does not rely on any parametric specification of the model (1.1) and is easy to implement. In the first stage, we estimate the conditional

distribution function of  $Y$  given  $X$  through nonparametric kernel methods. Secondly, we take a simple average of this estimate evaluated at the observed values of  $X^*$  to obtain an estimate of the CDF of  $Y^*$ . To see whether the counterfactual distribution differs from the original one, this result can be compared to an estimate of the distribution function of  $Y$ , such as the ordinary empirical CDF. Furthermore, any functional of the two CDFs can in turn be estimated by plugging in the respective estimators. For example, in order to obtain an estimate of the quantile function of  $Y^*$ , one can simply invert the estimate of the corresponding CDF.

We also provide a complete asymptotic theory for the estimation procedures proposed in this paper. A key result is that although our method is fully nonparametric, it is not affected by the curse of dimensionality: using empirical process theory, we show that our estimates of the functions of interest converge to certain Gaussian processes at the usual parametric rate  $\sqrt{n}$  irrespective of the dimension of  $X$ . We can therefore expect the asymptotics to provide a rather accurate approximation to the finite sample distribution even for moderate sample sizes. A further important result is that the ordinary nonparametric bootstrap works in our framework. This allows us to conduct asymptotically valid uniform inference on the entire functions being estimated, and not just some isolated points. We can thus test a number of important hypothesis involving the whole distribution of  $Y^*$  and  $Y$ , such as stochastic dominance ordering for example. The use of our methodology is illustrated through an empirical example and an extensive simulation study. The latter shows that our estimators and the related inferential procedures have good finite sample properties, even when the sample size is relatively small. Our approach should thus be appealing to applied researchers.

To the best of our knowledge, our paper is the first to consider estimation and inference for general distributional policy effects in a fully nonparametric framework. As such, it complements and extends an extensive literature on the estimation of policy effects in more restrictive settings. Stock (1989) and Imbens and Newey (2009) consider estimation of policy effects on the *mean* of the outcome variable in a nonparametric framework. DiNardo, Fortin, and Lemieux (1996), Donald, Green, and Paarsch (2000), Gosling, Machin, and Meghir (2000) and Machado and Mata (2005) develop policy estimators for more general distributional effects, but rely on various parametric restrictions of the model in (1.1). Furthermore, these papers generally focus on estimation rather than inference, and thus do not provide a full asymptotic theory for their procedures. Chernozhukov, Fernandez-Val, and Melly (2009) derive general limit distribution results for estimators of distributional policy effects, but again only for the case that model (1.1) is contained in certain parametric classes. In particular, their arguments critically rely on the

assumption that the conditional distribution function of  $Y$  given  $X$  can be estimated at a parametric rate, which is clearly not possible in our fully nonparametric framework. While using a correctly specified parametric model will obviously result in efficiency gains compared to our fully nonparametric procedure, such estimators will generally be inconsistent when the respective restrictions are violated. This trade-off is discussed in more detail as part of our simulation study.

In another related paper, Firpo, Fortin, and Lemieux (2009) propose a method to estimate the impact of a *marginal* increase in the covariates on the unconditional distribution of the outcome variable in a framework similar to ours. This parameter is different from the one being estimated in this paper, which corresponds to the effect of a general and *fixed* change in the distribution of the covariates. Furthermore, they use a very different estimation approach based on a linearization of the outcome distribution.

The outline of the paper is as follows. In the next section, we give a more formal description of our problem, define the parameters of interest, show under which conditions they can be identified, and describe the estimation procedure. Section 3 treats the asymptotic properties of our estimate of the distribution function, and Section 4 shows how these results can be used to analyse a wide range of statistics of the CDF. In Section 5 the practical relevance of our procedure is shown through a simulation study and a small-scale empirical application. Finally, Section 6 concludes. All proofs are collected in the Appendix.

## 2 Modelling Framework and Estimation Approach

### 2.1 Model

The setup we consider is as follows: we observe a dependent variable  $Y$  and a  $d$ -dimensional vector of covariates  $X$ , with marginal distribution functions  $F_Y$  and  $F_X$ , respectively. The dependent variable is assumed to be generated through the nonseparable model

$$Y = m(X, \varepsilon), \quad (2.1)$$

where  $\varepsilon$  is an unobserved error term. We assume that (2.1) is either a structural equation that describes the causal relationship between the right-hand and left-hand side variables, or a reduced form equation from a bigger structural system, as in Ichimura and Taber (2002). In a typical microeconometric application,  $X$  and  $\varepsilon$  would correspond to observed and unobserved characteristics of an individual, respectively, and  $m$  would describe the decision rule that, given

individual characteristics, determines the individual's choice  $Y$ . This flexible formulation allows the covariates to exert influence on  $Y$  in manifold ways. For example, model (2.1) allows for heteroskedasticity or skewness in the conditional distribution of  $Y$  given  $X$ . It is fully nonparametric in the sense that we do not restrict the function  $m$  or the distribution of the random variables involved to belong to some parametric family.

The values of at least some components of  $X$  are assumed to be under control of a (hypothetical) policy maker, and can thus be shifted exogenously to another observed random vector  $X^*$  with associated distribution function  $F_{X^*}$ . Substituting  $X^*$  for  $X$  in (2.1), we obtain the counterfactual random variable

$$Y^* = m(X^*, \varepsilon),$$

whose distribution function we denote by  $F_{Y^*}$ . Our interest is in learning (features of) this distribution and comparing it to that of  $Y$ .

For applications, the random vector  $X^*$  could originate from a number of different sources. First,  $X^*$  could be drawn from a different subpopulation corresponding to a different demographic group, geographic region or time period, like workers' characteristics in a different country for example. Second,  $X^*$  could be a deterministic transformation of  $X$ , i.e. there exists a known function  $\pi : \mathbb{R}^d \rightarrow \mathbb{R}^d$  such that  $X^* = \pi(X)$ . Examples of this case include a public program that causes smoking pregnant women to reduce their daily cigarette consumption by, say, half, or a tuition subsidy that is paid out subject to certain eligibility conditions. Third,  $X^*$  could be a repeated measurement on the same individual at a different point in time, as it would typically be the case when the data originate from a panel study.

While the specific source of  $X^*$  will not matter for our identification argument or the computation of the estimator, it is useful for the asymptotic development to distinguish those cases where  $X^*$  and  $X$  are stochastically independent, and those where they are not. We will call the former case an *independent policy implementation*, in the sense that the realization of  $X^*$  does not depend on the old value of  $X$ . The latter case is then accordingly termed as a *dependent policy implementation*.

## 2.2 Objects of Interest

Depending on the application at hand, a researcher might be interested in learning about different features of the distribution of  $Y^*$  and  $Y$ . Here we list some useful examples for which estimation and inference is discussed in detail below. However, our framework is by no means

limited to these examples, but can be used for any object that can be written as a sufficiently smooth (in the sense described below) functionals of the distribution functions of  $Y$  and  $Y^*$ .

Our primary objects of interest are the distribution functions  $F_Y$  and  $F_Y^*$  themselves. Assuming that the latter is identified, we could give a complete description of the effect of the policy on the distribution of the outcome variable by evaluating them directly. Furthermore, we will consider the pointwise difference between the two distribution functions,

$$\Delta_F(y) = F_Y^*(y) - F_Y(y), \quad (2.2)$$

which we call the Distribution Function Policy Effect. This quantity can be of interest in at least two respects. First, in some empirical contexts it might be sufficient to consider the effect of the policy on the distribution of the outcome variable at some fixed point only. In development economics for example, if  $Y$  is household income and  $\bar{y}$  is some fixed income level defined as the poverty line, one might be interested in whether the policy reduces the fraction of households that live under the poverty line, i.e. whether  $\Delta_F(\bar{y})$  is negative. Secondly, it might be interesting to test whether  $\Delta_F \equiv 0$ , and thus the policy has *any* kind of impact on the distribution of the outcome variable at all.

Instead of looking directly at the CDF, it is often more intuitive to consider the unconditional  $\tau$ -quantiles  $Q_Y^*$  and  $Q_Y$  of  $Y^*$  and  $Y$ , respectively, where

$$Q_Y^*(\tau) = \inf\{y \in \mathbb{R} : F_Y^*(y) \geq \tau\} \quad (2.3)$$

and  $Q_Y$  is defined analogously. Another convenient summary statistic is the corresponding Quantile Policy Effect, which is defined as

$$\Delta_Q(\tau) = Q_Y^*(\tau) - Q_Y(\tau). \quad (2.4)$$

This quantity is analogous to the quantile treatment effect in the literature on program evaluation.

In our framework, it is also possible to analyse the effect of the policy on the Lorenz curve and the Gini coefficient. These measures play an important role in the analysis of inequality and poverty. Formally, for a positive random variable with distribution function  $F_Y^*$  the Lorenz curve  $L_Y^*(p)$  is defined as the integral over the quantile function up to  $p$  divided by  $\mu$ , the mean of  $F$ :

$$L_Y^*(p) = \frac{1}{\mu_Y^*} \int_0^p Q_Y^*(\tau) d\tau = \frac{1}{\int_0^1 F_Y^{*-1}(\tau) d\tau} \int_0^p F_Y^{*-1}(\tau) d\tau, \quad (2.5)$$

The corresponding Gini coefficient is defined as twice the area between the Lorenz curve and the uniform distribution line, i.e.

$$G_Y^* = 1 - 2 \int_0^1 L_Y^*(p) dp, \quad (2.6)$$

with  $G = 0$  implying perfect equality and  $G = 1$  implying perfect inequality. The quantities  $L_Y$  and  $G_Y$  are then defined analogously. Again, we can also consider the Lorenz Curve Policy Effect, given by

$$\Delta_L(p) = L_Y^*(p) - L_Y(p), \quad (2.7)$$

and the Gini Policy Effect,

$$\Delta_G = G_Y^* - G_Y. \quad (2.8)$$

The final application we consider in this paper is testing for stochastic dominance. This topic is of great practical importance since the results can be used to evaluate the welfare implications of a proposed policy without making strong assumptions about social preferences. In particular, it is well known that if some distributions can be ranked by stochastic dominance, the same ranking is obtained through the corresponding social welfare over a wide range of utility functions (Atkinson 1970). To simplify the notation, define the operator  $\mathcal{D}_j(y, \phi)$  that integrates the function  $\phi$  up to order  $j - 1$  for  $j \geq 1$ , i.e.

$$\mathcal{D}_1(y, \phi) = \phi(y), \quad \mathcal{D}_2(y, \phi) = \int_0^y \phi(z) dz, \quad \mathcal{D}_3(y, \phi) = \int_0^y \int_0^t \phi(z) dz dt, \quad \text{etc.} \quad (2.9)$$

Then  $F_Y^*$  is said to dominate  $F_Y$  in " $j$ -th order stochastic dominance" sense if  $\mathcal{D}_j(y, F_Y^* - F_Y) \leq 0$  for all possible values of  $y$ . A Kolmogorov-Smirnov-type test for stochastic dominance can then be carried out by testing whether  $\sup_y \mathcal{D}_j(y, F_Y^* - F_Y)$  is negative.

### 2.3 Identification

For the setup considered in this paper, the only issue is whether the distribution functions of  $Y$  and  $Y^*$  are identified, since identification of the quantities (2.2)–(2.9) discussed in the previous subsection then follows trivially. While  $F_Y$  is obviously identified by the data, the case of  $F_Y^*$  is less clear. Identification in nonseparable models is a potentially delicate issue, that has attracted considerable interest in the recent literature (see for example Chesher (2003), Matzkin (2003), Imbens and Newey (2009) or Hoderlein and Mammen (2007)). However, since we are not interested in directly identifying features of  $m$ , our problem is much less complicated. In

particular, there is no need to impose anymore structure on the function  $m$ , such as monotonicity in the unobservables. For identification of our quantity of interest, the following assumption suffices.

**Assumption 1** (Identification). (i) *The term  $\varepsilon$  is independent of both  $X$  and  $X^*$ .* (ii) *The support of  $X^*$  is a subset of the support of  $X$ .*

**Proposition 1.** *Under Assumption 1,  $F_Y^*(y) = \mathbb{E}(F_{Y|X}(y, X^*))$  and is thus identified.*

The first part of Assumption 1 implies that  $X$  is exogenous, which is a strong assumption for many applications. It is straightforward to relax this condition by requiring independence to hold only conditional on some other random variable  $V$ , which is either observed or estimateable. Such a control variable could be available in a wide range of contexts, as described in Imbens and Newey (2009) for example. These include treatment effect models with selection on observables, triangular simultaneous equation models and certain sample selection models. In all cases, one can extend the result of Proposition 1 to  $F_Y^*(y) = \mathbb{E}(F_{Y|X,V}(y, X^*, V))$ . Our results on estimation given below then apply immediately when  $V$  is observed. The case where  $V$  has to be estimated (nonparametrically) from the data is technically much more involved and beyond the scope of this paper.

The second part of Assumption 1 restricts the policy experiments that can be considered to ones for which there is already some experience in the data. This restriction is due to the inability of nonparametric estimators to extrapolate from range of actual observations. While it limits the potential fields of application, without imposing some parametric structure on  $m$  this condition seems necessary to obtain point identification of  $F_Y^*$ . However, it is possible to give meaningful bounds on the CDF of  $Y^*$  when  $X^*$  is allowed to take values outside of the support of  $X$  with moderate probability.

## 2.4 Estimation

Our estimation approach is to first construct estimates  $\hat{F}_Y^*$  and  $\hat{F}_Y$  of the distribution functions  $F_Y^*$  and  $F_Y$ , respectively, and then estimate any functional of the form  $\Gamma = \Gamma(F_Y^*, F_Y)$  through the plug-in method by  $\hat{\Gamma} = \Gamma(\hat{F}_Y^*, \hat{F}_Y)$ . For example, an estimate of the Quantile Policy Effect  $\Delta_Q$  can be constructed as

$$\hat{\Delta}_Q(\tau) = \hat{Q}_Y^*(\tau) - \hat{Q}_Y(\tau) = \inf\{y \in \mathbb{R} : \hat{F}_Y^*(y) \geq \tau\} - \inf\{y \in \mathbb{R} : \hat{F}_Y(y) \geq \tau\}.$$

Estimates of all objects of interest defined in eq. (2.2)–(2.8) can be defined in an analogous manner.

The structure of the data used for the estimation depends on whether we are considering a *dependent* or an *independent policy implementation*. In the former case, the data consist of a sample  $\{(Y_i, X_i, X_i^*)\}_{i=1}^n$  of size  $n$  from the distribution of  $(Y, X, X^*)$ . For an independent policy implementation, the data consist of a sample  $\{(Y_i, X_i)\}_{i=1}^n$  of size  $n$  from the distribution of  $(Y, X)$ , and a sample  $\{X_i^*\}_{i=1}^{n^*}$  of size  $n^*$  from the distribution of  $X^*$ . While we allow the two samples sizes to be different in this case, we assume for the later asymptotic analysis that they increase proportionally, so that  $n^* = n/\lambda$  for some constant  $\lambda$ .

We now turn to the construction of the estimators. Throughout, we will use the notation for an independent policy implementation without loss of generality, since it covers the dependent implementation as the special case with  $n^* = n$ . Starting with an estimate for the CDF of  $Y$ , an obvious candidate is the usual empirical cumulative distribution function (ECDF),

$$\hat{F}_Y(y) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}\{Y_i \leq y\},$$

whose theoretical properties are well-known in the literature. To derive an estimator for  $F_Y^*$ , we know from the identification argument in the previous subsection that under Assumption 1

$$F_Y^*(y) = \mathbb{E}(F_{Y|X}(y, X^*)).$$

Following the analogy principle and replacing unknown quantities with suitable sample counterparts, it appears intuitive to use an estimator  $\hat{F}_Y^*$  of the form

$$\hat{F}_Y^*(y) = \frac{1}{n^*} \sum_{i=1}^{n^*} \hat{F}_{Y|X}(y, X_i^*),$$

where  $\hat{F}_{Y|X}$  is a first-stage nonparametric estimate of the distribution function of  $Y$  conditional on  $X$ . If all covariates are continuously distributed, we propose to estimate this function by a Nadaraya-Watson-type estimator, i.e.

$$\hat{F}_{Y|X}(y, x) = \frac{\hat{g}_{YX}(y, x)}{\hat{f}_X(x)}$$

where

$$\begin{aligned} \hat{g}_{YX}(y, x) &= \frac{1}{n} \sum_j \mathbb{I}\{Y_j \leq y\} K_{x,h}(X_j - x), \\ \hat{f}_X(x) &= \frac{1}{n} \sum_j K_{x,h}(X_j - x). \end{aligned}$$

Here  $\mathbb{I}\{A\}$  is an indicator function that equals one if  $A$  is true and zero otherwise,  $h = h_n$  is a bandwidth sequence that tends to zero as  $n \rightarrow \infty$ ,  $K_{x,h}(\cdot) = h^{-d}K_x(\cdot/h)$ , and  $K_x$  is a higher-order boundary kernel, i.e. a kernel function whose moments up to a certain order are zero,

and whose shape adapts when the point of evaluation  $x$  is in the vicinity of the boundary of the support of  $X$  (see for example Gasser, Müller, and Mammitzsch (1985)). These properties are needed to derive a uniform rate of convergence for our first-step estimator later, uniformly over the *entire* support of  $Y$  and  $X$ . We will be more precise about the specifics below.

The estimator  $\hat{F}_{Y|X}$  can easily be generalized to admit discrete regressors using the conventional frequency method. This entails splitting the sample into subsets, or cells, and then calculating the Nadaraya-Watson estimator within each such subset separately. This procedure is well known to have no effect on the rate of convergence. For notational convenience, we will therefore maintain the assumption that all covariates are continuously distributed.

A computational advantage of our estimator is that it admits a representation as a reweighted version of the usual empirical distribution function. To see this, note that  $\hat{F}_Y^*(y)$  can be written as

$$\begin{aligned}\hat{F}_Y^*(y) &= \frac{1}{n^*} \sum_{i=1}^{n^*} \frac{\sum_{j=1}^n \mathbb{I}\{Y_j \leq y\} K_{X_i^*, h}(X_j - X_i^*)}{\sum_{l=1}^n K_{X_i^*, h}(X_l - X_i^*)} \\ &= \frac{1}{n} \sum_{j=1}^n \mathbb{I}\{Y_j \leq y\} w_j\end{aligned}$$

where the weights  $w_j$  are given by

$$w_j = \lambda \sum_{i=1}^{n^*} \frac{K_{X_i^*, h}(X_j - X_i^*)}{\sum_{l=1}^n K_{X_i^*, h}(X_l - X_i^*)}.$$

Since the weights do not depend on  $y$ , they have to be calculated only once even when  $\hat{F}_Y^*$  is evaluated at multiple locations, making the estimator extremely cheap to compute in practice.

A potential caveat when using higher-order kernels is that they are not restricted to be positive, and hence can lead to estimates of  $F_Y^*$  which are non-monotone or take values outside the unit interval in finite samples, which is of course undesirable. This problem can be circumvented by a slight modification of the estimator: if we replace the weights  $w_j$  by  $\tilde{w}_j = w_j \mathbb{I}\{w_j \geq 0\} / \sum_i (w_i \mathbb{I}\{w_i \geq 0\})$ , we obtain a modified estimator

$$\tilde{F}_Y^*(y) = \frac{1}{n} \sum_{j=1}^n \mathbb{I}\{Y_j \leq y\} \tilde{w}_j$$

which is constrained to be monotonically increasing and bounded between 0 and 1. We show in the following section that this estimator asymptotically equivalent to  $\hat{F}_Y^*$  under standard conditions. For the further analysis, we will therefore assume without loss of generality that there are no issues with non-monotonicity of  $\hat{F}_Y^*$ .

### 3 Asymptotic Properties

In order to conduct inference on the CDFs and related functionals as a whole, we have to derive the joint asymptotic properties of the two estimators not only at some fixed value, but as a random function. To do so, we first state the assumptions and give some useful preliminary results in the following subsection, before proving the main weak convergence theorem in the next but one. Finally, we discuss inference and the validity of the bootstrap.

#### 3.1 Assumptions and Preliminaries

To present our framework, we first have to introduce some notation. For  $\mu$  a  $k$ -vector of nonnegative integers, we define (i)  $|\mu| = \sum_{i=1}^k \mu_i$ , (ii) for any function  $\phi(x)$  on  $\mathbb{R}^k$ ,  $\partial_x^\mu \phi(x) = \partial^{|\mu|}/(\partial^{\mu_1} x_1, \dots, \partial^{\mu_k} x_k) \phi(x)$  and (iii)  $x^\mu = \prod_{i=1}^k x_i^{\mu_i}$ . We write " $\xrightarrow{d}$ " to denote convergence in distribution of a sequence of random variables, and " $\Rightarrow$ " to denote weak convergence of a sequence of random functions.

To prove our main results, we need the following assumptions.

**Assumption 2.** *The data  $\{(Y_i, X_i)\}, i = 1, \dots, n$  and  $\{X_j^*\}, j = 1, \dots, n^*$  are i.i.d., respectively.*

**Assumption 3.** *(i) The support of  $X$  and  $X^*$  are the compact sets  $J = \otimes_{i=1}^d [x_i, \bar{x}_i]$  and  $J^* = \otimes_{i=1}^d [x_i^*, \bar{x}_i^*] \subset J$ , respectively. (ii)  $X$  has a probability density function  $f_X(x)$ , which is bounded away from zero on  $J$ . (iii)  $X^*$  has a probability density function  $f_X^*(x)$ , which is bounded away from zero on  $J^*$ . (iv) The functions  $f_X(x)$  and  $g(y, x) = F_{Y|X}(y, x)f_X(x)$  are  $r$ -times differentiable with respect to  $x$  on the interior of  $J$ , and the derivatives are uniformly continuous and bounded. (v) The function  $f_X^*(x)$  is  $r$ -times differentiable with respect to  $x$  on the interior of  $J^*$ , and the derivatives are uniformly continuous and bounded.*

**Assumption 4.** *Let  $D(c) = \{z : x - c \leq z \leq c - \bar{x}\}$ . Then the kernel function  $K_c : \mathbb{R}^d \rightarrow \mathbb{R}$  satisfies (i)  $\int_{D(c)} K_c(z) dz = 1$ , (ii)  $\int_{D(c)} K_c(z) z^\mu dz = 0$  for all  $|\mu| = 1, \dots, r$ , (iii)  $\int_{D(c)} |K_c(z)| z^\mu dz < \infty$  for  $|\mu| = r$ , (iv)  $K_c(z) = 0$  if  $|z| > 1$  (v)  $K_c(z)$  is  $r$ -times differentiable with respect to both  $z$  and  $c$ , and the derivatives are uniformly continuous and bounded.*

**Assumption 5.** *The bandwidth sequence  $h = h_n$  satisfies (i)  $h \rightarrow 0$ , (ii)  $n^{1/2}h^d/\log(n) \rightarrow \infty$  and (iii)  $n^{1/2}h^r \rightarrow 0$ .*

Assumption 2 is standard in microeconometric applications. Assumption 3 collects conventional smoothness restrictions on the functions being estimated through nonparametric methods

at some point in this paper. Note that it implicitly restricts the policies that can be considered to those where both  $X$  and  $X^*$  are continuously distributed. It is straightforward to show that this condition can be replaced by the assumption that both random vectors have a probability density function with respect to the same dominating measure. This would allow their components to be discrete or even continuous with some mass points, as long as the policy does not affect the location of the mass point. However, all policies for which  $X^*$  has a probability density function with respect to a different dominating measure than  $X$  are excluded in this framework. The fourth assumption prescribes the use of a higher-order boundary kernel, which is required to remove asymptotic bias from our first-step estimator. Note that the boundary correction is not necessary when  $J^*$  is a compact subset of the *interior* of  $J$ . Finally, the last assumption determines the rate at which the bandwidth sequence converges to zero. If  $h$  is of the form  $h = cn^{-\delta}$  for some constants  $c, \delta > 0$ , then in order for Assumption 6 to hold we need that  $\delta \in (1/2r, 1/2d)$  which in turn requires that the order of the kernel exceeds the dimension of  $X$ , so that the interval is not empty.

These assumptions are convenient, because they allow us to prove the following proposition, which is an important ingredient for the further arguments. Similar results have been obtained by Härdle, Janssen, and Serfling (1988) and Newey (1994a), to mention a few.

**Proposition 2.** *Under Assumption 1-5, we have that*

$$\begin{aligned} i) \quad & \sup_{y \in \mathbb{R}} \sup_{x \in J} |\hat{g}_{YX}(y, x) - g_{YX}(y, x)| = O_p \left( \left( \frac{\log n}{nh^d} \right)^{1/2} + h^r \right) \\ ii) \quad & \sup_{x \in J} |\hat{f}_X(x) - f_X(x)| = O_p \left( \left( \frac{\log n}{nh^d} \right)^{1/2} + h^r \right) \\ iii) \quad & \sup_{y \in \mathbb{R}} \sup_{x \in J} |\hat{F}_{Y|X}(y, x) - F_{Y|X}(y, x)| = O_p \left( \left( \frac{\log n}{nh^d} \right)^{1/2} + h^r \right) \end{aligned}$$

The proposition provides an explicit uniform rate of convergence for the first-step estimates. In particular, under Assumption 5 the proposition implies that the difference between  $\hat{F}_{Y|X}$  and  $F_{Y|X}$  vanishes at a rate faster than  $n^{-1/4}$ , whereas the corresponding bias disappears faster than  $n^{-1/2}$ , both uniformly over the two arguments. Thus our first-stage nonparametric estimator satisfies the well-known minimal convergence rates given by Newey (1994b).

Another important preliminary result is the asymptotic equivalence of the estimator  $\hat{F}_Y^*$  and its modified version  $\tilde{F}_Y^*$  introduced in Section 2.4. The following proposition implies that the limit results derived for  $\hat{F}_Y^*$  in the following section will also apply to  $\tilde{F}_Y^*$ , which has the practical advantage of being a proper distribution function.

**Proposition 3.** *Under Assumption 1-5, we have that*

$$\sup_{y \in \mathbf{R}} |\hat{F}_Y^*(y) - \tilde{F}_Y^*(y)| = o_p(n^{-1/2}).$$

### 3.2 Main Result

In this section, we derive the limit behaviour of our estimates of the distribution functions of  $Y^*$  and  $Y$ . In particular, we show that the bivariate random function

$$y \mapsto \sqrt{n} (\hat{\mathbf{F}}(y) - \mathbf{F}(y)) \tag{3.1}$$

converges weakly to some Gaussian process, where we use the notation that  $\hat{\mathbf{F}} = (\hat{F}_Y^*, \hat{F}_Y)^T$ ,  $\mathbf{F} = (F_Y^*, F_Y)^T$  and  $y = (y_1, y_2)^T$ .

The main complication for deriving this result originates from the process' first component, the normalized estimate of the CDF of  $Y^*$ , which is given by

$$\sqrt{n}(\hat{F}_Y^*(y_1) - F_Y^*(y_1)) = \sqrt{n} \left( \frac{1}{n^*} \sum_{i=1}^{n^*} \hat{F}_{Y|X}(y_1, X_i^*) - F_Y^*(y_1) \right).$$

The properties of this expression are not straightforward to derive, since our estimator  $\hat{F}_Y^*(y_1)$  is not a sum of independent terms:  $\hat{F}_{Y|X}(y_1, X_i^*)$  does not only depend on the  $i$ th but on all observations. In the appendix, we therefore construct an asymptotic representation for our estimate, which decomposes  $\hat{F}_Y^*$  into the following three parts:

$$\begin{aligned} \sqrt{n}(\hat{F}_Y^*(y_1) - F_Y^*(y_1)) &= \frac{1}{\sqrt{n^*}} \sum_{i=1}^{n^*} \sqrt{\lambda} (F_{Y|X}(y_1, X_i^*) - F_Y^*(y_1)) \\ &\quad + \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{f_X^*(X_i)}{f_X(X_i)} (\mathbb{I}\{Y_i \leq y_1\} - F_{Y|X}(y_1, X_i)) + o_p(1). \end{aligned}$$

The first term on the right hand side is the one we would obtain if the function  $F_{Y|X}$  was known and did not have to be estimated from the data.<sup>1</sup> It accounts for the uncertainty in our estimate that is induced by replacing the expectation with a sample average. The second term is an adjustment term that accounts for the uncertainty in our estimate of  $F_{Y|X}$ . The last term is  $o_p(1)$  uniformly in  $y$  and thus asymptotically negligible.<sup>2</sup> Using this decomposition and the definition of the ECDF, we arrive at the following Theorem:

---

<sup>1</sup>The factor  $\sqrt{\lambda}$  is an artefact of scaling the estimator by  $\sqrt{n}$  instead of  $\sqrt{n^*}$ .

<sup>2</sup>For the case of a fixed  $y$ , a similar decomposition for averages of nonparametrically estimated functions is shown by Newey (1994a).

**Theorem 1.** *If Assumptions 1-5 hold, then*

$$\sqrt{n} \left( \hat{\mathbf{F}}(\cdot) - \mathbf{F}(\cdot) \right) \Rightarrow \mathbb{F}_o(\cdot)$$

where  $\mathbb{F}_o$  is a two dimensional Gaussian process with mean zero and covariance function

$$\Psi^F(y, y') = \mathbb{E}(\psi^F(y, Z)\psi^F(y', Z)^T),$$

where  $Z = (Y, X, X^*)$  and  $\psi^F(y, Z) = (\psi_1^F(y_1, Z), \psi_2^F(y_2, Z))^T$  is given by

$$\psi_1^F(y_1, Z) = \sqrt{\lambda}(F_{Y|X}(y_1, X^*) - F_Y^*(y_1)) + \frac{f_X^*(X)}{f_X(X)}(\mathbb{I}\{Y \leq y_1\} - F_{Y|X}(y_1, X)),$$

$$\psi_2^F(y_2, Z) = \mathbb{I}\{Y \leq y_2\} - F_Y(y_2),$$

and the convergence is in  $D(-\infty, \infty) \times D(-\infty, \infty)$ .

The theorem shows that in large samples the behaviour of our random function (3.1) can be approximated by a bivariate correlated Gaussian process, whose second component is easily seen to be an ordinary  $F_Y$ -Brownian Bridge. The explicit form of the covariance function  $\Psi^F$  depends on whether we consider a dependent or independent policy implementation, but the influence function is the same in both cases. Although our estimator depends in part on high-dimensional nonparametric components, we obtain the  $\sqrt{n}$ -rate of convergence that one would typically obtain for standard parametric estimators, irrespective of the dimension of  $X$ . Our estimates are thus not affected by the curse of dimensionality, and hence we can expect the asymptotics to be a rather accurate approximation even when the sample size is only moderate relative to the dimension of  $X$ .

An immediate implication of Theorem 1 is that our estimator of the Distribution Function Policy Effect converges to a Gaussian process as well. That is, we obtain that

$$\sqrt{n}(\hat{\Delta}_F(\cdot) - \Delta_F(\cdot)) \Rightarrow (1, -1)\mathbb{F}_o(\cdot, \cdot).$$

by simply applying the continuous mapping theorem (CMT). In order to analyse the properties of estimates of more general functionals of the form  $\Gamma(\hat{\mathbf{F}})$ , one can use the following Theorem, which is an application of the Functional Delta Method (van der Vaart 2000, Theorem 20.8):

**Theorem 2.** *Suppose that the conditions of Theorem 1 hold, and let  $\Gamma$  be a Hadamard differentiable functional mapping from  $D(-\infty, \infty) \times D(-\infty, \infty)$  to some normed space  $S$ , with derivative  $\Gamma'_{\mathbf{F}}$ . Then*

$$\sqrt{n} \left( \Gamma(\hat{\mathbf{F}})(\cdot) - \Gamma(\mathbf{F})(\cdot) \right) \Rightarrow \Gamma'_{\mathbf{F}}(\mathbb{F}_o)(\cdot) \equiv \mathbb{G}_o(\cdot),$$

where  $\mathbb{G}_o$  is a Gaussian process with mean zero and covariance function

$$\Psi^\Gamma(y, y') = \mathbb{E}(\psi^\Gamma(y, Z)\psi^\Gamma(y', Z)^T)$$

with  $\psi^\Gamma(y, Z) = \Gamma'_F(\psi^F)(y, Z)$ , and the convergence is in  $S \times S$ .

The Hadamard differentiability condition in Theorem 2 requires the functional of interest to be sufficiently smooth around the true value  $\mathbf{F}$ . Roughly speaking, this means that  $\Gamma$  can locally be well approximated by some continuous linear functional  $\Gamma'_F$ , in the sense that

$$\frac{\Gamma(\mathbf{F} + ts_t) - \Gamma(\mathbf{F})}{t} \rightarrow \Gamma'_F(s) \quad \text{as } t \rightarrow 0$$

for all functions  $s_t \rightarrow s$  (see van der Vaart (2000, p. 296) for a precise definition). As we will see below, this condition can be verified for all our applications of interest under mild additional conditions. Also note that the theorem allows for functionals  $\Gamma$  that map into  $\mathbb{R}^k$  instead of some function space. In this case  $\mathbb{G}_o$  is simply a  $k$ -variate normal distribution, and  $\Psi^\Gamma(y, y') \equiv \Psi^\Gamma$  is its covariance matrix.

### 3.3 Inference

The results in Theorem 1 and 2 immediately provide the basis for conducting pointwise inference on certain features of the counterfactual distribution, by using the standard normal approximation. This might already be sufficient in some empirical contexts. In development economics for example, if  $Y$  is household income and  $\bar{y}$  is some fixed income level defined as the poverty line, one could simply be interested in whether the policy reduces the fraction of households that live under the poverty line, i.e. whether  $\Delta_F(\bar{y})$  is negative. Then it follows from the above results that the corresponding estimate  $\hat{\Delta}_F(\bar{y})$  is asymptotically normal with mean zero and standard error  $\sqrt{(1, -1)\Psi^F(\bar{y}, \bar{y})(1, -1)^T/n}$ , where  $\bar{y} = (\bar{y}, \bar{y})$ . Given a consistent estimate of the covariance function, one could thus test a null hypothesis such as  $H_0$ : "The policy does not change the proportion of people that earn below  $\bar{y}$ " through an ordinary  $t$ -statistic using standard normal critical values.

Many important hypotheses, however, cannot adequately be tested by considering only a fixed number of isolated points. This includes the hypothesis that the policy has no effect whatsoever, or that it leads to an improvement in a stochastic dominance sense. A related problem that involves the entire functions being estimated is the construction of uniform confidence bands, that cover the true function with some prespecified probability.

One possibility to address these problems would be to simulate the limiting processes in Theorem 1 and 2 by so-called multiplier methods (see e.g. van der Vaart and Wellner (1996, Section 2.9)). A disadvantage of this approach is that it requires explicit calculation and consistent estimation of the covariance function, which can be a cumbersome task for some applications. A natural alternative to using simulation methods is to conduct inference using a form of the bootstrap, for which one does not necessarily need to be able to give an explicit characterization of the limiting distribution of the process of interest. In particular, by using the bootstrap one can circumvent explicit specification of the covariance function.

In this paper, we propose using a simple nonparametric (or empirical) bootstrap scheme, which is based on resampling the original observations. For the implementation, we again have to take possible dependencies between  $X$  and  $X^*$  into account. For a dependent policy implementation, the bootstrap data is given by a sample  $\{(Y_{b,i}, X_{b,i}, X_{b,i}^*)\}_{i=1}^n$  drawn with replacement from  $\{(Y_i, X_i, X_i^*)\}_{i=1}^n$ , whereas for an independent policy implementation it consists of two samples  $\{(Y_{b,i}, X_{b,i})\}_{i=1}^n$  and  $\{X_{b,i}^*\}_{i=1}^{n^*}$  drawn with replacement from  $\{(Y_i, X_i)\}_{i=1}^n$  and  $\{X_i^*\}_{i=1}^{n^*}$ , respectively. In both cases, the resampled data is then used to calculate the bootstrap estimates of  $F_Y^*$  and  $F_Y$ , which are denoted by  $\hat{\mathbf{F}}_b = (\hat{F}_{Y,b}^*, \hat{F}_{Y,b})$ , using the estimator described in Section 2.4. The distribution of  $\hat{\mathbf{F}}_b$  can then be determined through the usual repeated resampling of the data, and used as an approximation of the distribution of  $\hat{\mathbf{F}}$ . The following Theorem gives a theoretical justification for this approach.

**Theorem 3.** *Suppose that the conditions of Theorem 1 hold. Then*

$$\sqrt{n} \left( \hat{\mathbf{F}}_b(\cdot) - \hat{\mathbf{F}}(\cdot) \right) \Rightarrow \mathbb{F}_o(\cdot),$$

*conditional on the data, in probability, and the convergence is in  $D(-\infty, \infty) \times D(-\infty, \infty)$ .*

*Furthermore, under the conditions of Theorem 2,*

$$\sqrt{n} \left( \Gamma(\hat{\mathbf{F}}_b)(\cdot) - \Gamma(\hat{\mathbf{F}})(\cdot) \right) \Rightarrow \Gamma'_{\mathbf{F}}(\mathbb{F}_o)(\cdot),$$

*conditional on the data, in probability<sup>3</sup>, and the convergence is in  $S \times S$ .*

Theorem 3 states that the nonparametric bootstrap is not only valid for the original problem of conducting inference on the estimates of the CDFs directly, but that it can also be used to analyse the properties of general functionals of the CDFs as well. As a simple example, consider

---

<sup>3</sup>See van der Vaart and Wellner (1996, Section 3.9.3) for a precise definition of conditional weak convergence in probability.

the problem of forming a uniform  $1 - \alpha$  confidence band for the Distribution Function Policy Effect  $\Delta_F(\cdot)$ . To this end, let  $\hat{\Delta}_{F,b}(y) = \hat{F}_{Y,b}^*(y) - \hat{F}_{Y,b}(y)$  the bootstrapped Distribution Function Policy Effect, and define the pointwise variance of  $\hat{\Delta}_{F,b}$  with respect to bootstrap sampling as  $\sigma^2(y) = \text{Var}_b(\hat{\Delta}_{F,b}(y))$ . It then follows directly from Theorem 3 that a uniform  $1 - \alpha$  confidence band for  $\Delta_F$  is given by

$$CB_{1-\alpha}(y) = [\hat{\Delta}_F(y) - c\sigma(y), \hat{\Delta}_F(y) + c\sigma(y)],$$

where  $c$  is the smallest positive constant that satisfies

$$P_b \left( \sup_y \left| (\hat{\Delta}_{F,b}(y) - \hat{\Delta}_F(y)) \right| / \sigma(y) \leq c \right) \geq 1 - \alpha \quad (3.2)$$

and  $P_b$  is the probability with respect to bootstrap sampling. In practice, the unknown quantities  $c$  and  $\sigma^2(\cdot)$  can be approximated with the usual resampling techniques.

## 4 Application to Objects of Interest

In this section, we use results from Theorem 1 – 3 do discuss the remaining applications of interest from Section 2.2. We show that in each case plug-in type estimators converge to a Gaussian limit, and that the nonparametric bootstrap can be used to conduct asymptotically valid inference, under appropriate additional regularity conditions.

### 4.1 Quantiles

To analyse the properties of the estimators of quantiles and Quantile Policy Effect, we use the fact that inversion operator that transforms a CDF into its corresponding quantile function is a Hadamard differentiable functional, which gives us the following proposition:

**Proposition 4.** *Assume that (i)  $F_Y^*$  and  $F_Y$  are both continuously differentiable with strictly positive derivative  $f_Y^*$  and  $f_Y$ , and (ii)  $Y^*$  and  $Y$  have compact support, then*

$$\sqrt{n}(\hat{\mathbf{Q}}(\cdot) - \mathbf{Q}(\cdot)) \Rightarrow - \left( \frac{\mathbb{F}_{o1}}{f_Y^*}, \frac{\mathbb{F}_{o2}}{f_Y} \right) \circ \mathbf{Q} \equiv \mathbb{Q}_o$$

and

$$\sqrt{n}(\hat{\mathbf{Q}}_b(\cdot) - \hat{\mathbf{Q}}(\cdot)) \Rightarrow \mathbb{Q}_o(\cdot)$$

conditional on the data, in probability. Here  $\mathbb{Q}_o$  is a Gaussian process with mean zero and covariance function

$$\Psi^Q(\tau, \tau') = \mathbb{E}(\psi^Q(\tau, Z)\psi^Q(\tau', Z)^T)$$

where

$$\psi^Q(\tau, Z) = \left( \frac{\psi_1^F(Q_Y^*(\tau_1), Z)}{f_Y^*(Q_Y^*(\tau_1))}, \frac{\psi_2^F(Q_Y(\tau_2), Z)}{f_Y(Q_Y(\tau_2))} \right)^T.$$

and the convergence is in  $\ell^\infty(0, 1) \times \ell^\infty(0, 1)$ .

We thus have the familiar result that the influence function of the quantile process is just the influence function of the corresponding distribution function divided by the density of the variable of interest, and evaluated at the respective quantile. The assumption that  $Y^*$  and  $Y$  have compact support can be relaxed at the cost of restricting convergence to subsets of the unit interval. Furthermore, it follows directly from the proposition and the continuous mapping theorem that our estimator of the Quantile Policy Effect satisfies

$$\sqrt{n}(\hat{\Delta}_Q(\cdot) - \Delta_Q(\cdot)) \Rightarrow (1, -1)\mathbb{Q}_o(\cdot),$$

and that the nonparametric bootstrap is valid in this case as well. Thus, one can construct uniform confidence bands on the Quantile Policy Effect in exactly the same manner as for the Distribution Function Policy Effect.

## 4.2 Inequality measures

In this section, we apply Theorem 1–3 to analyse the effect of a proposed policy on the Lorenz curve and the Gini coefficient. Our approach is similar to that of Barrett and Donald (2009) and Bhattacharya (2007), who in different contexts also obtained weak convergence results using the functional delta method. Many other inequality measures such as the Theil index for example can be treated by analogous arguments. We start by deriving the asymptotic properties of the estimated Lorenz curves.

**Proposition 5.** *Assume that (i)  $F_Y^*$  and  $F_Y$  are both continuously differentiable with derivatives  $f_Y^*$  and  $f_Y$ , respectively, (ii) these derivatives are strictly positive on any compact subset of  $(0, \infty)$ , (iii)  $Y^*$  and  $Y$  have finite second moments, and (iv) it holds that*

$$\lim_{y \rightarrow \infty} \frac{(1 - F_Y(y))^{1+c}}{f_Y(y)} = \lim_{y \rightarrow 0} \frac{F_Y(y)^{1+c}}{f_Y(y)} = 0$$

for some  $0 < c < 1$ , and similarly for  $Y^*$ . Furthermore, define the process  $\mathbb{H}_o$  as

$$\mathbb{H}_o(p) = \int_0^p \mathbb{Q}_o(\tau) d\tau$$

for  $p = (p_1, p_2)^T$ , and the integral is understood to be taken componentwise. Then

$$\sqrt{n} \left( \hat{\mathbf{L}}(\cdot) - \mathbf{L}(\cdot) \right) \Rightarrow \left( \frac{\mathbb{H}_{o1}(\cdot)}{\mu_Y^*} - \frac{L_Y^*(\cdot)}{\mu_Y^*} \mathbb{H}_{o1}(1), \frac{\mathbb{H}_{o2}(\cdot)}{\mu_Y} - \frac{L_Y(\cdot)}{\mu_Y} \mathbb{H}_{o2}(1) \right) \equiv \mathbb{L}_o(\cdot),$$

and

$$\sqrt{n} \left( \hat{\mathbf{L}}_{\mathbf{b}}(\cdot) - \hat{\mathbf{L}}(\cdot) \right) \Rightarrow \mathbb{L}_o(\cdot),$$

conditional on the data, in probability. Here  $\mathbb{L}_o$  is a Gaussian process with mean zero and covariance function

$$\Psi^L(p, p') = \mathbb{E}(\psi^L(p, Z) \psi^L(p', Z)^T)$$

where  $\psi^L(p, Z) = (\psi_1^L(p_1, Z), \psi_2^L(p_2, Z))^T$  is given in Appendix B, and the convergence is in  $\ell^\infty(0, 1) \times \ell^\infty(0, 1)$ .

The additional assumptions we make are used by Bhattacharya (2007) in order to establish Hadamard-differentiability of the functional that translates a CDF into its Lorenz curve. Note that in contrast to the quantile process, one does not have to assume that the support of  $Y$  and  $Y^*$  is compact, and thus that their density functions are bounded away from zero, to obtain weak convergence in  $\ell^\infty(0, 1) \times \ell^\infty(0, 1)$ . Instead, the tail condition (iv), which is fulfilled by many distributions commonly used to describe income distributions such as log-normal or Pareto, suffices.

Proposition 4 could again be used to conduct inference on the Lorenz curve as a whole. Also, as in the case of the quantiles considered above, it follows from the continuous mapping theorem that our estimator of the Lorenz Policy Effect satisfies

$$\sqrt{n}(\hat{\Delta}_L(\cdot) - \Delta_L(\cdot)) \Rightarrow (1, -1)\mathbb{L}_o(\cdot),$$

and that the nonparametric bootstrap is valid in this case as well. Here the bootstrap could be used for example to construct a one-sided confidence band  $\hat{\Delta}_L$ , which could be used to test the hypothesis of Lorenz dominance, i.e. that  $\Delta_L(p) \leq 0$  for all  $p \in (0, 1)$ . Another direct consequence of the proposition is the distribution of the Gini coefficient, which again follows simply from the continuous mapping theorem.

**Proposition 6.** *Under the same conditions as Proposition 4,*

$$\sqrt{n}(\hat{\mathbf{G}} - \mathbf{G}) \xrightarrow{d} N(0, \Psi^G) \quad \text{and} \quad \sqrt{n}(\hat{\mathbf{G}}_{\mathbf{b}} - \hat{\mathbf{G}}) \xrightarrow{d} N(0, \Psi^G),$$

conditional on the sample, in probability, where the limiting distribution is bivariate normal with mean zero and covariance matrix

$$\Psi^G = 4\mathbb{E} \left( \int_0^1 \psi^L(p, Z) dp \int_0^1 \psi^L(p, Z)^T dp \right).$$

### 4.3 Testing for Stochastic Dominance

The limit results from Section 3 can also be used for various testing problems. Here we adapt the methods of Barrett and Donald (2003) and consider Kolmogorov-Smirnov-type statistics to test stochastic dominance of  $F_Y^*$  over  $F_Y$  for any prespecified order, with critical values obtained via the bootstrap. Other approaches are discussed, for example, by McFadden (1989), Anderson (1996), Davidson and Duclos (2000), Abadie (2002) and Linton, Maasoumi, and Whang (2005).

The test statistics we consider are based on pointwise comparisons of appropriate measures of distance between  $\hat{F}_Y^*$  and  $\hat{F}_Y$  over their entire common support, which we assume to be the compact interval  $[0, \bar{y}]$ <sup>4</sup>. Using the operator  $\mathcal{D}_j(y, \phi)$  defined in (2.9), the hypothesis of "j-th order stochastic dominance" of  $F_Y^*$  over  $F_Y$  can then be formulated as

$$\begin{aligned} H_0^j : \quad & \mathcal{D}_j(y, F_Y^* - F_Y) \leq 0 \quad \forall y \in [0, \bar{y}], \\ H_1^j : \quad & \mathcal{D}_j(y, F_Y^* - F_Y) > 0 \quad \exists y \in [0, \bar{y}], \end{aligned}$$

for  $j = 1, 2, \dots$ . The corresponding test statistics are given by

$$KS_j = \sqrt{n} \sup_{y \in [0, \bar{y}]} \mathcal{D}_j(y, \hat{F}_Y^* - \hat{F}_Y) = \sqrt{n} \max_{y \in Y_1, \dots, Y_n} \mathcal{D}_j(y, \hat{F}_Y^* - \hat{F}_Y),$$

where the last equality follows from the fact that by construction both  $\hat{F}_Y^*$  and  $\hat{F}_Y$  are piecewise constant functions that jump at observed values of  $Y$  only.

Our aim is to reject  $H_0^j$  whenever  $KS_j$  exceeds some critical value. Note that there can be many different combinations of distribution functions  $F_Y^*$  and  $F_Y$  such that  $H_0^j$  is true, and we are thus testing a composite null hypothesis. However, it is easy to see that the least favourable case in this context corresponds to  $F_Y^* = F_Y$ . An asymptotically valid test procedure can thus be based on bootstrapping the test statistic under the least favourable null. In particular, one can calculate the bootstrap  $p$ -value as

$$\hat{p}_j = B^{-1} \sum_{b=1}^B \mathbb{I}\{\hat{KS}_{j,b} > KS_j\},$$

---

<sup>4</sup>Focussing on random variables that take only positive values seems natural, since stochastic dominance tests are usually applied to income or wealth distributions. The upper limit on the support is needed for the proofs and is usually not restrictive for empirical applications.

where

$$\hat{KS}_{j,b} = \sqrt{n} \max_{y \in Y_1, \dots, Y_n} \left( \mathcal{D}_j(y, \hat{F}_{Y,b}^* - \hat{F}_{Y,b}) - \mathcal{D}_j(y, \hat{F}_Y^* - \hat{F}_Y) \right)$$

is the realisation of the test statistic when calculated from the bootstrap sample. Our test decision can then be based on the following rule:

$$\text{Reject } H_0^j \text{ if } \hat{p}_j < \alpha \text{ for some prespecified significance level } \alpha. \quad (4.1)$$

The following proposition delivers a theoretical justification for this approach.

**Proposition 7.** *For any  $j = 0, 1, 2, \dots$  and  $\alpha < 1/2$ , the decision rule (4.1) is a test of  $H_0^j$  vs.  $H_1^j$  that has (i) asymptotic size of at most  $\alpha$  and (ii) is consistent against any fixed alternative.*

## 5 Numerical Examples

### 5.1 Simulation Study

#### 5.1.1 Setup

In order to demonstrate the usefulness of our proposed estimation procedures, we conduct a number of simulation experiments to assess their finite sample properties. Specifically, we simulate a vector  $X = (X_1, X_2, X_3)$  of conditioning variables, where the three components are i.i.d. standard exponentially distributed and truncated at 3, and generate the dependent variable of interest  $Y$  through a linear model with conditional heteroskedasticity as

$$Y = 6 - 2X_1 + X_2 + \sigma(X)\varepsilon, \quad \sigma^2(X) = X_1 + X_2, \quad (5.1)$$

and  $\varepsilon$  follows a standard exponential distribution. Note that  $Y$  is restricted to be positive in this setup, and that  $X_3$  is an irrelevant regressor. We consider two dependent policy implementations, where  $X^*$  is a deterministic transformation of the original  $X$ -values:

- *Policy 1:*  $X_3$  is reduced by 50%:  $\pi_1(x_1, x_2, x_3) = (x_1, x_2, .5x_3)$ .
- *Policy 2:* All regressors are reduced by 50%:  $\pi_2(x_1, x_2, x_3) = (.5x_1, .5x_2, x_3)$ .

Since  $X_3$  does not appear in the data generating process of  $Y$ , the first policy has no effect on the distribution of  $Y$ . In contrast, Policy 2 highly affects the dependent variable, leading to a distribution of  $Y^*$  that second-order stochastically dominates that of  $Y$ . The corresponding distribution functions  $F_Y$  and  $F_Y^*$  are plotted in Figure 5.1.

Figure 5.1: Plot of  $F_Y$  (solid line) and  $F_Y^*$  (dashed line) for Policy 2.

For both policies, we consider the applications of interest described in Section 2.2. Here the CDFs of  $Y^*$  and  $Y$ , and the Distribution Function Policy Effect  $\Delta_F$  are estimated over the equidistant grid  $\{3, 3.05, 3.1, \dots, 8.95, 9\}$ , whereas for the quantiles of  $Y^*$  and  $Y$ , the Quantile Policy Effect, the Lorenz curve of  $Y^*$  and  $Y$  and the Lorenz Curve Policy Effect the grid  $\{.1, .11, .12, \dots, .89, .9\}$  is used. Moreover, we consider tests for first- and second-order stochastic dominance of  $F_Y^*$  over  $F_Y$ . That is, in each simulation run we test the hypothesis  $H_0^j : \mathcal{D}_j(y, F_Y^* - F_Y) \leq 0$  for  $j = 1, 2$ . We use the sample sizes  $n = 100$  and  $n = 400$ , and set the number of Monte Carlo replications to 1000. In each replication, we use the nonparametric bootstrap with  $B = 1000$  repetitions to obtain uniform 90% confidence bands for the functionals of interest (and approximate  $p$ -values in case of the stochastic dominance tests).

In order to implement our estimators, we have to specify a kernel function and a bandwidth sequence  $h$  that are compatible with the assumptions made in Section 3. A kernel function that

satisfies both the higher-order and the boundary correction property is given by

$$K_c(z) = \prod_{i=1}^d e_1^T S_{c_i}^{-1}(1, z_i, \dots, z_i^p)^T \kappa(z_i),$$

where  $S_c = (\mu_{j+l,c})_{0 \leq j,l \leq p}$  is a matrix of kernel constants  $\mu_{j,c} = \int_{D(c)} z^j \kappa(z) dz$ ,  $e_1 = (1, 0, \dots, 0)^T$  is the unity vector,  $p = r - 3$  and  $\kappa(z)$  is a standard univariate kernel function that satisfies the remaining regularity conditions of Assumption 4. This is the product of univariate equivalent kernels of a local polynomial regression estimator (see Fan and Gijbels (1996) for more details). For our simulations, we let  $\kappa(z)$  be the usual Epanechnikov kernel and choose  $p = 1$ , which implies that  $K_c(z)$  is a fourth order kernel.

Regarding the bandwidth, our asymptotic results only prescribe a rate at which  $h$  tends to zero, but are silent about its optimal size in finite samples. Here we use a bandwidth of the form  $h = cn^{-\delta}$ , which requires that  $\delta \in (1/2r, 1/2d) = (1/8, 1/6)$  in order for Assumption 5 to be fulfilled. Absent further guidelines, we choose  $h = 1.5\sigma_x n^{-1/7}$  for our simulations, where  $\sigma_x$  is the standard deviation of the respective covariates. Informal robustness checks suggest that the results are not too sensitive with respect to this choice.

### 5.1.2 Results

In Table 1, we present the result of our simulation study regarding the properties of our estimates of the CDFs of  $Y^*$  and  $Y$ , the Distribution Function Policy Effect, the quantiles of  $Y^*$  and  $Y$ , the Quantile Policy Effect, the Lorenz curve of  $Y^*$  and  $Y$ , the Lorenz curve policy effect, the Gini coefficients of  $Y^*$  and  $Y$  and the Gini Policy effect. In each case, we report Monte Carlo estimates of the integrated bias (IBias), the root integrated mean squared error (RIMSE), and the coverage rate of a uniform confidence band with nominal coverage level of 90% (Cov. Rate).

Although the sample sizes we consider are relatively small, our estimators exhibit reasonable finite sample properties. Finite sample biases are generally small and decrease rapidly with the sample size. Also note that increasing the sample size from  $n = 100$  to  $n = 400$ , i.e. by a factor of four, roughly halves the magnitude of the RIMSE for all quantities under consideration, which indicates that convergence to the true values indeed takes places at rate  $\sqrt{n}$ . The empirical coverage rate of the uniform 90% bootstrap confidence bands is generally close to the nominal level. This procedure should thus be able to provide reliable inference even in small samples.

Table 2 presents the simulation results on the stochastic dominance tests. For each test, we report the empirical rejection rates for the nominal levels  $\alpha = .05$  and  $\alpha = .1$ . Recall that for the first policy  $F_Y^* \equiv F_Y$  so that  $H_0^j$  is true for  $j = 1, 2$ , whereas under Policy 2  $F_Y^*$  is

Table 1: Simulation Results: Properties of Nonparametric Policy Estimators

	<i>Policy 1</i>			<i>Policy 2</i>		
	<i>n</i> = 100			<i>n</i> = 400		
	IBias	RIMSE	CR	IBias	RIMSE	CR
$F_Y^*$	0.879	9.469	0.855	0.414	5.358	0.894
$F_Y$	0.654	7.926	0.813	0.349	3.975	0.862
$\Delta_F$	0.367	5.059	0.952	0.277	3.553	0.964
$Q_Y^*$	4.495	30.421	0.900	1.000	16.063	0.898
$Q_Y$	2.143	22.777	0.843	0.626	11.611	0.861
$\Delta_Q$	3.557	21.288	0.973	0.453	11.185	0.941
$L_Y^*$	0.150	1.008	0.833	0.066	0.542	0.877
$L_Y$	0.135	0.842	0.805	0.049	0.435	0.846
$\Delta_L$	0.015	0.566	0.972	0.016	0.340	0.964
$G_Y^*$	0.353	1.770	0.885	0.154	0.950	0.901
$G_Y$	0.310	1.493	0.860	0.115	0.769	0.871
$\Delta_G$	0.042	0.954	0.955	0.039	0.596	0.947
	<i>n</i> = 100			<i>n</i> = 400		
	IBias	RIMSE	CR	IBias	RIMSE	CR
$F_Y^*$	2.169	10.531	0.862	0.899	5.211	0.857
$F_Y$	0.266	7.698	0.847	0.326	3.978	0.865
$\Delta_F$	2.109	7.513	0.859	0.699	3.717	0.876
$Q_Y^*$	2.698	20.428	0.868	1.098	9.877	0.861
$Q_Y$	0.580	22.265	0.875	0.591	11.466	0.853
$\Delta_Q$	2.833	21.669	0.948	0.811	10.731	0.920
$L_Y^*$	0.063	0.668	0.825	0.016	0.337	0.856
$L_Y$	0.014	0.843	0.812	0.012	0.429	0.865
$\Delta_L$	0.059	0.780	0.855	0.012	0.379	0.878
$G_Y^*$	0.098	1.150	0.877	0.007	0.586	0.873
$G_Y$	0.028	1.483	0.879	0.026	0.762	0.870
$\Delta_G$	0.126	1.361	0.864	0.019	0.664	0.885

Note: Integrated Bias and RIMSE figures have been multiplied by 100 to improve readability.

Table 2: Simulation Results: Rejection rates of KS-type tests for stochastic dominance

		<i>Policy 1</i>		<i>Policy 2</i>	
		$KS_1$	$KS_2$	$KS_1$	$KS_2$
$\alpha = .05$	$n = 100$	.036	.040	.345	.011
	$n = 400$	.048	.041	.917	.012
$\alpha = .1$	$n = 100$	.074	.083	.538	.025
	$n = 400$	.098	.089	.983	.011

dominating  $F_Y$  in a second-order stochastic dominance sense, but not in a first-order one, so that only  $H_0^2$  holds in this case. For Policy 1, both tests are conservative, but their empirical size gets closer to the respective nominal level as the sample size increases. Under Policy 2, the  $KS_1$  test has non-trivial power for  $n = 100$  and rejects the null in almost all simulation runs for  $n = 400$ . The rejection rates of the  $KS_2$  test are substantially below their nominal values, which comes as no surprise as the test can only be expected to have correct size under the least favourable null hypothesis.

### 5.1.3 Comparison with Approach based on Quantile Regression

Without any point of reference, it is admittedly difficult to judge whether the finite sample properties of our estimators are "good". In this section, we therefore briefly compare them with those of an estimator based on a first-step linear quantile regression (LQR), as discussed in Machado and Mata (2005), Melly (2005) and Chernozhukov, Fernandez-Val, and Melly (2009). Instead of using a kernel estimator, this method obtains an estimate of the conditional distribution function  $F_{Y|X}$  by inverting an estimate of the conditional quantile function  $Q_{Y|X}$ , which is assumed to be linear in the regressors at every quantile, i.e. it imposes the restriction that  $Q_{Y|X}(\tau|x) = x\beta(\tau)$  for all  $\tau \in (0, 1)$ . Since this estimator imposes additional parametric restrictions on the relationship between the dependent variable and the covariates, the result will generally exhibit less finite sample variation than our nonparametric procedure. On the other hand, such an estimator is also more prone to misspecification bias. The purpose of this section is to illustrate this tradeoff.

We compare the properties of the LQR-based estimator to our nonparametric procedure via simulation. For brevity, we restrict attention on the Quantile Policy Effect, and consider only the effect of Policy 2 for  $n = 400$ . The setup we use is the same as described above, with the

Table 3: Simulation results: Comparisson of estimators.

	$a = 0$			$a = .5$			$a = 1$		
	IBias	RIMSE	CR	IBias	RIMSE	CR	IBias	RIMSE	CR
NP	1.197	10.623	.920	2.202	13.972	.940	3.363	18.249	.934
LQR	1.883	9.319	.917	8.051	15.082	.702	16.743	26.074	.294

Note: Integrated Bias and RIMSE figures have been multiplied by 100 to improve readability.

exception that the dependent variable is now generated as

$$Y = 6 - 2X_1 + X_2 + aX_2^2 + \sigma(X)\varepsilon, \quad \sigma^2(X) = X_1 + X_2. \quad (5.2)$$

The parameter  $a$  governs the complexity of the relationship. For  $a = 0$ , (5.2) is the same as (5.1) considered above. In this case, a LQR model would be correctly specified. To illustrate the effect of misspecification, we also consider  $a = .5$  and  $a = 1$ .

The result of the simulations, given in Table 3, show that our nonparametric estimator compares favourably with its competitor and performs well uniformly over the different values of  $a$  we considered. It has the lowest RIMSE under all designs except for  $a = 0$ , where it exceeds the RIMSE of the correctly specified LQR-based estimator by about 15%. When the underlying model is not correctly specified, the LQR-based estimator can exhibit a substantial bias, with its magnitude depending on the degree of misspecification. The coverage rates of uniform confidence bands can also deviate significantly from their nominal levels in this case.

## 5.2 Empirical Illustration: The Effect of an Anti-Smoking Campaign on Infant Birthweights

In this section, we illustrate the application of our estimators through an example form public health. We consider a (hypothetical) public policy that successfully induces women who smoke during their pregnancy to cut their average daily cigarette consumption by 75%. Our interest is in the effect of this policy on the distribution of infant birthweight in general, and whether it helps to reduce the incidence of low-birthweight infants, which is usually defined by infants weight at birth falling below 2500 grams (about 5 pounds, 8 ounces). These issues should be of concern to policy makers since low birthweight is known to be associated with a wide range of subsequent health problems, and has even been linked to later educational attainment and labor market outcomes (see for example Almond, Chay, and Lee (2005) or Black, Devereux, and Salvanes (2007)).

The data we use is a subsample of the Detailed Natality Data (June 1997) published by the

Table 4: Descriptive Statistics

	Mean	Std. Dev	Min	Q25	Median	Q75	Max
Birthweight	3176	560.36	457	2889	3204	3515	5245
Cigarettes per day	11.98	7.51	1	6.5	10	20	60
Mother's age	25.15	5.56	18	21	24	29	45
Mother's Weight Gain	30.19	14.07	0	20	30	40	98
Married	0.52	0.49	0	—	—	—	1

Note:  $N = 4439$ . Married is an indicator variable for the mother being married.

National Center for Health Statistics. A more extensive analysis of the full data set is given by Abrevaya (2001) and Koenker and Hallock (2001). The subsample we employ comprises 4439 white mothers between age 18 and 45 without any college education, who gave birth to a live, single infant and smoked during their pregnancy. For each woman in this particular subgroup, we record the infant's birthweight (in grams) and the average daily number of cigarettes the mother smoked during the pregnancy, together with other variables that could possibly confound the relationship between birthweight and the level of cigarette consumption. These include the mother's age, mother's weight gain during the pregnancy (in pounds) and whether the mother is married. Table 4 presents some descriptive statistics for our full data set. The identifying assumption is that conditional on these variables unobserved factors that influence birthweight are independent of the average amount of cigarettes consumed, given that the mother chooses to smoke during the pregnancy in the first place.

Using the same specifications as for our simulation study in Section 5.1, we estimate the effect of reducing every woman's cigarette consumption by 75% on the distribution of birthweights by our nonparametric procedure. The discrete regressor is accommodated using the conventional frequency method. Figure 5.2 presents the estimate of the Quantile Policy Effect  $\Delta_Q(\cdot)$ , together with 90% confidence intervals for every point, and a 90% uniform confidence band, both obtained via the bootstrap with  $B = 1000$  replications. The graph suggest that the policy increases infant birthweights over almost the entire range of quantiles considered. Moreover, the point estimate suggests that particularly the low quantiles would benefit by such a campaign, with an estimated increase in the 10% quantile of about 140 grams compared to only 70 grams at the 90% quantile. However, since the associated confidence band is relatively wide, one cannot reject the hypothesis that  $\Delta_Q(\cdot)$  is constant, and thus there is no significant evidence of heterogeneous policy effects.

As mentioned above, another quantity of interest is the proportion of low-weight birth

Figure 5.2: Estimated Quantile Policy Effect on Infant Birthweight (solid line), with pointwise 90% confidence bands (dark-grey area) and uniform 90% confidence bands (light-grey area) based on 1000 bootstrap replications.

incidents, which amounts to 9.28% in the subpopulation under consideration. To see how a change in smoking habit would affect this share, we simply estimate the Distribution Function Policy Effect  $\Delta_F(\cdot)$  and evaluate it 2500 grams. As a result, we obtain that

$$\hat{\Delta}_F(2500) = -0.0290 \quad \text{with} \quad s.e.(\hat{\Delta}_F(2500)) = 0.0092.$$

The corresponding 90% confidence interval is given by  $CI_{.9} = (-0.0429, -0.0140)$ . This implies that the policy would reduce low-weight birth incidents by roughly one third, which is a substantial amount. While the confidence interval is again relatively wide, it is substantially to the left of zero, which indicates that the policy should be effective for reducing low-birthweight incidents.

## 6 Conclusions

In this paper, we have proposed a fully nonparametric way to assess the effect of an exogenous change in the distribution of the covariates on the unconditional distribution of a dependent variable of interest. The method can be used to conduct asymptotically valid inference on various kinds of statistics that can be written as a sufficiently smooth functional of the CDF. It is straightforward to implement and performs surprisingly well in simulations when the sample size is relatively small, even compared to correctly specified parametric estimators. This is in sharp contrast to classical nonparametric methods, whose use is prohibitive due to the curse of dimensionality in situations with many regressors and few observations.

## A Proofs of Theorems and Propositions

### A.1 Proof of Proposition 1

From the definition of a distribution function, the law of iterated expectations, and that  $\varepsilon$  is independent of both  $X$  and  $X^*$ , we obtain that

$$\begin{aligned} F_Y^*(y) &= P(m(X^*, \varepsilon) \leq y) \\ &= \int P(m(X^*, \varepsilon) \leq y | X^* = x) dF_X^*(x) \\ &= \int P(m(x, \varepsilon) \leq y) dF_X^*(x) \\ &= \int P(m(X, \varepsilon) \leq y | X = x) dF_X^*(x) \\ &= \mathbb{E}(F_{Y|X}(y, X^*)), \end{aligned}$$

where  $F_{Y|X}$  is the conditional distribution function of  $Y$  given  $X$ . Since the data  $(Y, X)$  certainly identify this function at any point  $(y, x)$  in their support, and  $X^*$  only takes values in a subset of the support of  $X$  with probability 1, this implies that  $F_Y^*$  is identified.  $\square$

### A.2 Proof of Proposition 2

The statement can be shown using standard kernel smoothing theory. First, one can show that the rate of the bias of each estimator is uniformly of the order  $O(h^r)$  by standard Taylor expansion arguments. The fact that the rate is the same in both the interior and the vicinity of the boundary of the support of  $X$  is a consequence of the use of a boundary kernel. Second, the rate of the stochastic part can be shown to be  $O_p(\log(n)/(nh^d))$  by using the arguments from, say, Newey (1994a) for (ii), and Härdle, Janssen, and Serfling (1988) or Akritas and Van Keilegom (2001) for (i) and (iii). Taken together, these results imply the rates given in the proposition.  $\square$

### A.3 Proof of Proposition 3

Consider for notational simplicity the case that  $\lambda = 1$ , and define the function  $w$  through

$$w(x) = \frac{1}{n} \sum_{i=1}^{n^*} \frac{K_{X_i^*, h}(x - X_i^*)}{\hat{f}_X(X_i^*)},$$

so that  $w_j = w(X_j)$ . Noting that  $\hat{F}_Y^* \equiv \tilde{F}_Y^*$  when  $w(x)$  is positive for every value of  $x$ , we obtain that

$$\begin{aligned} P(\hat{F}_Y^* = \tilde{F}_Y^*) &\geq P(\inf_{x \in J} w(x) \geq 0) \\ &\geq P(\inf_{x \in J} \hat{f}_X(x) \geq c \cap \inf_{x \in J} \hat{f}_X^*(x) \geq 0) \\ &\geq P(\inf_{x \in J} \hat{f}_X(x) \geq c) + P(\inf_{x \in J} \hat{f}_X^*(x) \geq 0) - 1 \end{aligned} \quad (\text{A.1})$$

for some constant  $c > 0$ , uniformly in  $y$ . We now show that both probabilities in (A.1) are of the order  $1 + o_p(n^{-1/2})$  if  $c$  is chosen sufficiently small, which gives the desired result.

To see this for the first term in (A.1), note that the density of  $X$  is bounded away from zero on  $J$ , i.e. there exists some  $\delta > 0$  such that  $\inf_{x \in J} f_X(x) > \delta$ . Now, set  $c = \delta/2$ . It is a well-known result that  $\mathbb{E}(\hat{f}_X(x)) = f_X(x) + O(h^r)$  uniformly in  $x$ , and we thus know that  $\inf_x \mathbb{E}(\hat{f}_X(x)) > \frac{3}{4}\delta$  for some sufficiently large  $n$ . This also means that for sufficiently large  $n$  the event  $\{\inf_x \hat{f}_X(x) \geq c\}$  is implied by the event  $\{\inf_x (\hat{f}_X(x) - \mathbb{E}(\hat{f}_X(x))) \geq -c/2\}$ , which is in turn implied by the event  $\{\sup_x |\hat{f}_X(x) - \mathbb{E}(\hat{f}_X(x))| < c/2\}$ . That is, it holds that

$$P(\inf_{x \in J} \hat{f}_X(x) \geq c) \geq P(\sup_{x \in J} |\hat{f}_X(x) - \mathbb{E}(\hat{f}_X(x))| < c/2).$$

Now, since  $J$  is compact, it can be covered by  $v_n \leq \gamma_1 n^d$  open balls with radius  $n^{-1}$ , for some  $\gamma_1 > 0$ . The  $k$ th of these balls, with midpoint  $x_{n,k}$ , is denoted by

$$J_{n,k} = \{x \in \mathbb{R}^d : \|x - x_{n,k}\| \leq n^{-1}\}.$$

Then we have that

$$\sup_{x \in J} |\hat{f}_X(x) - \mathbb{E}(\hat{f}_X(x))| \leq \max_{1 \leq k \leq v_n} \sup_{x \in J_{n,k}} |\hat{f}_X(x) - \mathbb{E}(\hat{f}_X(x))|.$$

For  $x \in J_{n,k}$ , it follows from the triangle inequality that

$$|\hat{f}_X(x) - \mathbb{E}(\hat{f}_X(x))| \leq |\hat{f}_X(x) - \hat{f}_X(x_{n,k})| + |\hat{f}_X(x_{n,k}) - \mathbb{E}(\hat{f}_X(x_{n,k}))| + |\mathbb{E}(\hat{f}_X(x_{n,k})) - \mathbb{E}(\hat{f}_X(x))|.$$

Since the kernel function  $K$  has bounded partial derivatives, the first and third term on the right-hand side of the last equation can be bounded as follows:

$$\begin{aligned} |\hat{f}_X(x) - \hat{f}_X(x_{n,k})| &\leq C \frac{|x - x_{n,k}|}{h^{d+1}} \leq C n^{-1} h^{-d-1}, \\ |\mathbb{E}(\hat{f}_X(x_{n,k})) - \mathbb{E}(\hat{f}_X(x))| &\leq C \frac{|x - x_{n,k}|}{h^{d+1}} \leq C n^{-1} h^{-d-1} \end{aligned}$$

for some  $C > 0$ . Thus

$$\sup_{x \in J} |\hat{f}_X(x) - \mathbb{E}(\hat{f}_X(x))| \leq \max_{1 \leq k \leq v_n} |\hat{f}_X(x_{n,k}) - \mathbb{E}(\hat{f}_X(x_{n,k}))| + 2Cn^{-1}h^{-d-1}. \quad (\text{A.2})$$

Using a result from Bosq (1998, p.47–48), we also have that

$$\begin{aligned} P\left(\max_{1 \leq k \leq v_n} |\hat{f}_X(x_{n,k}) - \mathbb{E}(\hat{f}_X(x_{n,k}))| > c/4\right) &\leq \sum_{k=1}^{v_n} P(|\hat{f}_X(x_{n,k}) - \mathbb{E}(\hat{f}_X(x_{n,k}))| > c/4) \\ &\leq v_n C \exp(-\gamma_2 \sqrt{nh^d}) \\ &\leq C \exp(-\gamma_2 \sqrt{nh^d} + \gamma_1 \log(n)) \end{aligned} \quad (\text{A.3})$$

for some  $\gamma_2 > 0$ . Since  $n^{-1}h^{-d-1} \rightarrow 0$  by Assumption 5, (A.2) and (A.3) together imply that for  $n$  sufficiently large

$$\begin{aligned} P\left(\sup_{x \in J} |\hat{f}_X(x) - \mathbb{E}(\hat{f}_X(x))| < c/2\right) &\geq P\left(\max_{1 \leq k \leq v_n} |\hat{f}_X(x_{n,k}) - \mathbb{E}(\hat{f}_X(x_{n,k}))| < c/4\right) \\ &\geq 1 - C \exp(-\gamma_2 \sqrt{nh^d} + \gamma_1 \log(n)) \\ &= 1 + o_p(n^{-1/2}) \end{aligned}$$

if  $nh^d/\log(n)^2 \rightarrow \infty$ , which is again ensured through Assumption 5. An analogous argument then applies to the second term in (A.1). This completes our proof.  $\square$

#### A.4 Proof of Theorem 1

In this section, we will briefly switch to the operator notation typically used in the empirical process literature. In particular, for  $A_1, \dots, A_n$  an i.i.d. sequence of random variables taking values in  $(\mathcal{A}, \mathcal{B})$  with distribution  $P$  and for some measurable function  $\phi : \mathcal{X} \rightarrow \mathbb{R}$  we will write

$$P\phi = \int \phi dP, \quad \mathbb{P}_n\phi = \frac{1}{n} \sum_i \phi(Z_i), \quad \mathbb{G}_n\phi = \sqrt{n}(\mathbb{P}_n - P)\phi$$

for the expectation, empirical measure and empirical process at  $\phi$ , respectively. Furthermore, we write  $\bar{o}_p(a_n)$  as a shorthand notation for " $o_p(a_n)$  uniformly in  $y \in \mathbb{R}$ ".

The difficulties in deriving the limit behaviour of  $\mathbb{F}_n$  arise from the fact that  $\hat{F}_Y^*$  is itself a random function that has been estimated from the data. Using the notation described above, the first component of  $\mathbb{F}_n$  can be rewritten as

$$\sqrt{n}(\hat{F}_Y^* - F_Y^*) = \sqrt{n}(\mathbb{P}_{n^*}\hat{F}_{Y|X} - P^*F_{Y|X})$$

where  $P^*$  is the distribution of  $X^*$  and both  $\hat{F}_{Y|X}$  and  $F_{Y|X}$  are seen as functions of  $X^*$  indexed by  $y$ , i.e.  $X^* \mapsto \hat{F}_{Y|X}(y, X^*)$  and similarly for  $F_{Y|X}$ . The above expression can be decomposed into the following three terms,

$$\begin{aligned} \sqrt{n}(\mathbb{P}_{n^*}\hat{F}_{Y|X} - P^*F_{Y|X}) &= \sqrt{\lambda}\mathbb{G}_{n^*}^*(\hat{F}_{Y|X} - F_{Y|X}) + \sqrt{\lambda}\mathbb{G}_{n^*}^*F_{Y|X} + \sqrt{n}P^*(\hat{F}_{Y|X} - F_{Y|X}) \\ &= T_1 + T_2 + T_3, \end{aligned}$$

which can now be analyzed separately. Beginning with  $T_1$ , it is shown below in Lemma 1 that

$$\sup_{y \in \mathbb{R}} \left| \mathbb{G}_{n^*}^*(\hat{F}_{Y|X} - F_{Y|X}) \right| = o_p(1)$$

and thus the first term on the right hand side vanishes as  $n$  tends to infinity. The second term is equal to

$$T_2 = \frac{1}{\sqrt{n^*}} \sum_{i=1}^{n^*} \sqrt{\lambda}(F_{Y|X}(y, X_i^*) - F_Y^*(y)),$$

which does not contain any unknown functions and is thus easy to handle. Finally, Lemma 2 establishes that

$$T_3 = \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{f_X^*(X_i)}{f_X(X_i)} (\mathbb{I}\{Y_i \leq y\} - F_{Y|X}(y, X_i)) + o_p(1).$$

To derive the asymptotic distribution for the actual process of interest, we have to distinguish between a dependent and an independent policy implementation, and we have to introduce some more notation. First, define the functions

$$\begin{aligned} \xi_{11}(y) &: X^* \mapsto \sqrt{\lambda} F_{Y|X}(y, X^*) \\ \xi_{12}(y) &: X^* \mapsto 0 \\ \xi_{21}(y) &: (Y, X) \mapsto \frac{f_X^*(X)}{f_X(X)} (\mathbb{I}\{Y \leq y\} - F_{Y|X}(y, X)) \\ \xi_{22}(y) &: (Y, X) \mapsto \mathbb{I}\{Y \leq y\} - F_Y(y) \end{aligned}$$

and let  $\xi_1 = (\xi_{11}, \xi_{12})^T$  and  $\xi_2 = (\xi_{21}, \xi_{22})^T$  and  $\xi = \xi_1 + \xi_2$ . Second, define the classes of functions  $\mathcal{P}_{ij} = \{\xi_{ij}(y); y \in \mathbb{R}\}$  for  $i, j = 1, 2$  and  $\mathcal{P}_i = \{\xi_i(y); y \in \mathbb{R}^2\}$  for  $i = 1, 2$  and  $\mathcal{P} = \{\xi(y); y \in \mathbb{R}^2\}$ . Then it can be shown that each of these classes of functions is Donsker by combining the results in the Examples 19.6, 19.9 and 19.20 in van der Vaart (2000), and by noting that if two classes of functions are Donsker then so is their Cartesian product (van der Vaart 2000, p. 270).

Now consider the case of an independent policy implementation. Letting  $P$  be the distribution of  $(Y, X)$ , the process of interest can be written as

$$\sqrt{n} (\hat{\mathbf{F}} - \mathbf{F}) = \mathbb{G}_{n^*}^* \xi_1 + \mathbb{G}_n \xi_2 + o_p(1).$$

Since  $\mathcal{P}_1$  and  $\mathcal{P}_2$  are Donsker, the first and second term on the right hand side converge to two independent Gaussian process. Thus, by the continuous mapping theorem, the entire right hand side of the last equation converges to the sum of these two Gaussian processes, which is again a Gaussian process by independence. It is then straightforward to see that this has mean zero and covariance function as stated in the Theorem.

For a dependent policy implementation, let  $\bar{P}$  be the joint distribution of  $(Y, X, X^*)$  and recall that  $\lambda = 1$  in this case. The process can then be written as

$$\sqrt{n} (\hat{\mathbf{F}} - \mathbf{F}) = \bar{\mathbb{G}}_n \xi + o_p(1).$$

which converges to a Gaussian process since  $\mathcal{P}$  is Donsker. It is again straightforward to see that the limiting process has again mean zero and covariance function as stated in the Theorem (with  $\lambda = 1$ ). This completes our proof.  $\square$

We now prove the two lemmas used in the above argument.

**Lemma 1.** *Under the conditions of Theorem 1, it holds that*

$$\sup_{y \in \mathbb{R}} \left| \mathbb{G}_{n^*}^*(\hat{F}_{Y|X} - F_{Y|X}) \right| = o_p(1)$$

*Proof.* By Lemma 19.24 in van der Vaart (2000), the statement of the lemma follows if (i) the sequence of random functions  $\hat{F}_{Y|X}$  takes its values in some Donsker class  $\mathcal{F}$ , and if (ii)

$$\sup_{y \in \mathbb{R}} P^*(\hat{F}_{Y|X} - F_{Y|X})^2 = o_p(1).$$

This last condition obviously holds in our case because

$$\begin{aligned} P^*(\hat{F}_{Y|X} - F_{Y|X})^2 &= \int (\hat{F}_{Y|X}(y, x) - F_{Y|X}(y, x))^2 dF_X^*(x) \\ &= o_p(n^{-1/2}) \end{aligned}$$

uniformly in  $y$  and  $x$  since  $\|\hat{F}_{Y|X}(y, x) - F_{Y|X}(y, x)\|_\infty = o_p(n^{-1/4})$  by Proposition 2, and the integration takes place over a compact set.

To see that the first condition holds, define the class  $\mathcal{F}$  by

$$\mathcal{F} = \{\hat{F}_{Y|X}(y, \cdot) : y \in \mathbb{R}\},$$

and note that by Assumption 4 each of its elements is  $r$  times continuously differentiable, and the derivatives are uniformly bounded. Hence  $\mathcal{F} \subset C_M^r(J^*)$ , the space of all functions on  $J^*$ , the support of  $X^*$ , whose derivatives up to order  $r$  are uniformly bounded by some constant  $M$ . But this class is Donsker if  $r > d/2$ , where  $d$  is dimension of  $J^*$ , as shown in van der Vaart (2000, Example 19.9). However, our assumptions on the bandwidth already require that  $r > d$ , so that this condition easily holds in our case.  $\square$

**Lemma 2.** *Under the conditions of Theorem 1,*

$$\sqrt{n}P^*(\hat{F}_{Y|X} - F_{Y|X}) = \frac{1}{\sqrt{n}} \sum_{i=1}^n (\mathbb{I}\{Y_i \leq y\} - F_{Y|X}(y, X_i)) \frac{f_X^*(X_i)}{f_X(X_i)} + o_p(1)$$

*uniformly in  $y$ .*

*Proof.* Switching back to the usual notation, we have that

$$\sqrt{n}P^*(\hat{F}_{Y|X} - F_{Y|X}) = \sqrt{n} \int \hat{F}_{Y|X}(y, x) - F_{Y|X}(y, x) dF_X^*(x).$$

Here and in the following, all integrals are understood to be taken over the entire  $\mathbb{R}^d$ . In order to derive an expression for the integral in the above equation, it is useful to split it up into two components.

$$\begin{aligned}
& \int \hat{F}_{Y|X}(y, x) - F_{Y|X}(y, x) dF_X^*(x) \\
&= \int \frac{1}{n} \sum_i (\mathbb{I}\{Y_i \leq y\} - F_{Y|X}(y, X_i)) \frac{K_{x,h}(x - X_i)}{\hat{f}_X(x)} dF_X^*(x) \\
&\quad + \int \frac{1}{n} \sum_i (F_{Y|X}(y, X_i) - F_{Y|X}(y, x)) \frac{K_{x,h}(x - X_i)}{\hat{f}_X(x)} dF_X^*(x) \\
&= A + B
\end{aligned}$$

We start with analysing the term  $A$ . Using the fact that  $X$  is a continuous random variable with density function  $f_X$ , and applying a second order Taylor expansion of  $1/\hat{f}_X(x)$  around  $1/f_X(x)$ , we obtain

$$\begin{aligned}
A &= \int \frac{1}{n} \sum_i (\mathbb{I}\{Y_i \leq y\} - F_{Y|X}(y, X_i)) \frac{K_{x,h}(x - X_i)}{\hat{f}_X(x)} f_X^*(x) dx \\
&= \int \frac{1}{n} \sum_i (\mathbb{I}\{Y_i \leq y\} - F_{Y|X}(y, X_i)) \frac{K_{x,h}(x - X_i)}{f_X(x)} f_X^*(x) dx \\
&\quad - \int \frac{1}{n} \sum_i (\mathbb{I}\{Y_i \leq y\} - F_{Y|X}(y, X_i)) (\hat{f}_X(x) - f_X(x)) \frac{K_{x,h}(x - X_i)}{f_X(x)^2} f_X^*(x) dx \\
&\quad + o_p(n^{-1/2}) \\
&= A_1 + A_2 + o_p(n^{-1/2})
\end{aligned}$$

where the last term is  $o_p(n^{-1/2})$  uniformly in  $x$  and  $y$  since  $\|\hat{f}_X(x) - f_X(x)\|_\infty = o_p(n^{-1/4})$ ,  $|\mathbb{I}\{Y_i < y\} - F_{Y|X}(y, X_i)| \leq 1$ , and integration takes place over a compact set.

To derive an expression for  $A_1$ , define  $u(x) = f_X^*(x)/f_X(x)$ , let  $u^{(\mu)}(x) = \partial_x^\mu u(x)$  and  $K_c^{(\mu)}(x) = \partial_c^\mu K_c(x)$ . Using standard techniques from the kernel smoothing literature, we obtain that

$$\begin{aligned}
A_1 &= \int \frac{1}{n} \sum_i (\mathbb{I}\{Y_i \leq y\} - F_{Y|X}(y, X_i)) K_{x,h}(x - X_i) u(x) dx \\
&= \frac{1}{n} \sum_i (\mathbb{I}\{Y_i \leq y\} - F_{Y|X}(y, X_i)) \int K_{zh+X_i}(z) u(zh + X_i) dz \\
&= \frac{1}{n} \sum_i (\mathbb{I}\{Y_i \leq y\} - F_{Y|X}(y, X_i)) \int (K_{X_i}(z) + zh K_{X_i}^{(1)}(z) + \dots + (zh)^r K_\xi^{(r)}(z)) \\
&\quad \times (u(X_i) + zh u^{(1)}(X_i) + \dots + (zh)^r u^{(r)}(\xi)) dz \\
&= \frac{1}{n} \sum_i (\mathbb{I}\{Y_i \leq y\} - F_{Y|X}(y, X_i)) (u(X_i) + O_p(h^r)) \\
&= \frac{1}{n} \sum_i (\mathbb{I}\{Y_i \leq y\} - F_{Y|X}(y, X_i)) \frac{f^*(X_i)}{f(X_i)} + o_p(n^{-1/2})
\end{aligned}$$

where  $\xi$  is some value between  $X_i$  and  $X_i + zh$ . Here the second-to-last equality follows from interchanging the order of differentiation and integration (which in turn follows by dominated convergence) and using the kernel properties, and the last equality holds because  $O_p(h^r) = o_p(n^{-1/2})$  by Assumption 5.

Next, we consider the term  $A_2$ . Plugging in the definition of  $\hat{f}_X$ , we obtain

$$\begin{aligned}
A_2 &= \int \frac{1}{n} \sum_i (\mathbb{I}\{Y_i \leq y\} - F_{Y|X}(y, X_i)) (\hat{f}(x) - f(x)) \frac{K_h(x - X_i)}{f_X(x)^2} f_X^*(x) dx \\
&= \frac{1}{n} \sum_i (\mathbb{I}\{Y_i \leq y\} - F_{Y|X}(y, X_i)) \int \left( \frac{1}{n} \sum_j K_{x,h}(X_j - x) - f(x) \right) K_{x,h}(x - X_i) \frac{f_X^*(x)}{f_X(x)^2} dx \\
&= \frac{1}{n^2} \sum_{i,j} (\mathbb{I}\{Y_i \leq y\} - F_{Y|X}(y, X_i)) \int (K_{x,h}(X_j - x) - f(x)) K_{x,h}(x - X_i) \frac{f_X^*(x)}{f_X(x)^2} dx \\
&= \frac{1}{n^2} \sum_{i,j} (\mathbb{I}\{Y_i \leq y\} - F_{Y|X}(y, X_i)) \\
&\quad \times \left( \int K_{x,h}(X_j - x) K_{x,h}(x - X_i) \frac{f_X^*(x)}{f_X(x)^2} dx - \int K_{x,h}(x - X_i) \frac{f_X^*(x)}{f_X(x)^2} dx \right) \\
&= \frac{1}{n^2} \sum_{i,j} (\mathbb{I}\{Y_i \leq y\} - F_{Y|X}(y, X_i)) (A_{21} - A_{22})
\end{aligned}$$

Applying the same kind of argument as for the derivation of  $A_1$ , we obtain

$$A_{22} = \frac{f_X^*(X_i)}{f_X(X_i)} + o_p(n^{-1/2})$$

Turning to the first term, defining  $v(x) = f_X^*(x)/f_X^2(x)$ , using similar arguments as before, we obtain that

$$\begin{aligned}
A_{21} &= \int v(x) K_{x,h}(X_j - x) K_{x,h}(x - X_i) dx \\
&= \int v(zh + X_i) K_{zh+X_i,h}(X_j - X_i - zh) K_{zh+X_i}(z) dz \\
&= v(X_i) K_{X_i,h}(X_j - X_i) + \bar{o}_p(n^{-1/2}).
\end{aligned}$$

Hence we have shown that

$$\begin{aligned}
A_2 &= \frac{1}{n^2} \sum_{i,j} v(X_i) (\mathbb{I}\{Y_i \leq y\} - F_{Y|X}(y, X_i)) (K_{X_i,h}(X_j - X_i) - f_X(X_i)) + \bar{o}_p(n^{-1/2}) \\
&= U_n(y) + \bar{o}_p(n^{-1/2}).
\end{aligned}$$

To proceed, we need to introduce some further notation. Let  $\bar{f}_X(x) = \mathbb{E}(K_{x,h}(X_j - x))$ , and define  $H(Z_i, Z_j; y, h) = v(X_i) (\mathbb{I}\{Y_i \leq y\} - F_{Y|X}(y, X_i)) (K_{X_i,h}(X_j - X_i) - \bar{f}_X(X_i))$ . Then, because by Proposition 2 we have that  $\bar{f}_X(x) - f_X(x) = o_p(n^{-1/2})$  uniformly in  $x$ , we can write  $U_n(y)$  as

$$U_n(y) = \frac{1}{n^2} \sum_i \sum_{j \neq i} H(Z_i, Z_j; y, h) + \frac{1}{n^2} \sum_i H(Z_i, Z_i; y, h) + o_p(n^{-1/2}). \quad (\text{A.4})$$

It is straightforward to see that the first term on the right hand side of (A.4) is a degenerate second order U-process. It then follows from Corollary 4 in Sherman (1994), a Uniform Law of Large Numbers for U-processes, that

$$\sup_{y \in \mathbb{R}} \left| \frac{1}{n^2} \sum_i \sum_{j \neq i} H(Z_i, Z_j; y, h) \right| = O_p(h^{-d} n^{-1}).$$

To see that the conditions of Sherman (1994, Corollary 4) are satisfied, let

$$\mathcal{H} = \{h^d H(\cdot; y, h), y \in \mathbb{R}, h > 0\}.$$

Since  $h^d H(\cdot; y, h)$  is uniformly bounded as a function of both  $y$  and  $h$ , the class  $\mathcal{H}$  has a constant (and hence square integrable) envelope function. The class also satisfies the so-called euclidean property required for Corollary 4, by Assumption 3 and 4, Lemma 22(ii) in Nolan and Pollard (1987) and Lemma 2.14 of Pakes and Pollard (1989). Hence the conditions of the corollary are satisfied.

One can furthermore directly see that the second term in (A.4) is also  $O_p(h^{-d}n^{-1})$  uniformly in  $y$ , and thus  $A_{22} = \bar{o}_p(n^{-1/2})$  because  $O_p((nh^d)^{-1}) = o_p(n^{-1/2})$  by Assumption 5. This completes our argument regarding the term  $A$ . Turning to the term  $B$ , it can be shown that  $B = o_p(n^{-1/2})$  uniformly in  $y$  through similar reasoning.  $\square$

## A.5 Proof of Theorem 2

The statement of the Theorem follows directly from the Functional Delta Method (see Section 3.9 in van der Vaart and Wellner (1996))  $\square$

## A.6 Proof of Theorem 3

To see the first assertion of the theorem, consider the case of a dependent policy implementation. Recall from the proof of Theorem 1 that

$$\sqrt{n} (\hat{\mathbf{F}} - \mathbf{F}) = \bar{\mathbb{G}}_n \xi + o_p(1).$$

Since  $\mathcal{P} = \{\xi(y); y \in \mathbb{R}^2\}$  is Donsker, the result then follows from Theorem 3.6.1 in van der Vaart and Wellner (1996). An analogous argument can be made for the case of an independent policy implementation. The second assertion of the theorem is then simply a consequence of the Function Delta Method for the bootstrap, see Theorem 3.9.11 in van der Vaart and Wellner (1996).  $\square$

## A.7 Proof of Proposition 3

By Lemma 21.4 in van der Vaart (2000), under the conditions of the proposition the map  $\Gamma$  with

$$\Gamma(\phi) = (\phi_1^{-1}, \phi_2^{-1}) \quad \text{and } \phi_i^{-1}(\tau) = \inf\{y : \phi_i(y) \geq \tau\}, \quad i = 1, 2$$

is Hadamard differentiable at  $\mathbf{F}$  tangentially to  $C(0, 1) \times C(0, 1)$ , with derivative

$$\phi \mapsto \Gamma'_{\mathbf{F}}(\phi) = - \left( \frac{\phi_1}{f_Y^*}, \frac{\phi_2}{f_Y} \right) \circ \mathbf{F}^{-1}.$$

Thus, by Theorem 2 the joint quantile process  $\sqrt{n}(\hat{\mathbf{Q}} - \mathbf{Q})$  converges weakly to  $\mathbb{Q}_o$  in  $\ell^\infty(0, 1) \times \ell^\infty(0, 1)$ . The validity of the bootstrap then follows directly from Theorem 3.  $\square$

## A.8 Proof of Proposition 4

Our process of interest  $\sqrt{n}(\hat{\mathbf{L}} - \mathbf{L})$  can be rewritten as

$$\sqrt{n}(\hat{\mathbf{L}} - \mathbf{L}) = \sqrt{n}(\Gamma_L(\hat{\mathbf{F}}) - \Gamma_L(\mathbf{F}))$$

with the "Lorenz functional"  $\Gamma_L$  defined as

$$\begin{aligned}\Gamma_L(\phi)(p) &= \left( \int_0^{p_1} \phi_1^{-1}(\tau) d\tau, \int_0^{p_2} \phi_2^{-1}(\tau) d\tau \right) \cdot \left( \frac{1}{\int_0^\infty \phi_1^{-1}(\tau) d\tau}, \frac{1}{\int_0^\infty \phi_2^{-1}(\tau) d\tau} \right) \\ &\equiv S(\phi)(p) \cdot \mu^{-1}(\phi)\end{aligned}$$

for  $p = (p_1, p_2)^T$ . Using this notation, we can rewrite the Lorenz process as

$$\sqrt{n}(\hat{\mathbf{L}} - \mathbf{L}) = \mu^{-1}(\hat{\mathbf{F}}) \cdot \sqrt{n}(S(\hat{\mathbf{F}}) - S(\mathbf{F})) - \mu^{-1}(\hat{\mathbf{F}}) \cdot \mathbf{L} \cdot \sqrt{n}(\mu(\hat{\mathbf{F}}) - \mu(\mathbf{F}))$$

The asymptotic properties of this expression can now be derived by looking at the individual components. First, since  $\mu^{-1}(\cdot)$  is a continuous functional, it follows from Theorem 1 and the continuous mapping theorem that

$$\mu^{-1}(\hat{\mathbf{F}}) \rightarrow \mu^{-1}(\mathbf{F}) \equiv (1/\mu_Y^*, 1/\mu_Y),$$

where  $\mu_Y^*$  and  $\mu_Y$  are the unconditional means of  $Y^*$  and  $Y$ , respectively. Second, using a result from Bhattacharya (2007, Claim 1), it follows that the map  $S$  is Hadamard differentiable at  $\mathbf{F}$  tangentially to  $C[0, 1] \times C[0, 1]$  with derivative

$$\phi \mapsto S'_{\mathbf{F}}(\phi)(p) = \left( \int_0^{p_1} \frac{\phi_1(F_Y^{*-1}(\tau))}{f_Y^*(F_Y^{*-1}(\tau))} d\tau, \int_0^{p_2} \frac{\phi_2(F_Y^{-1}(\tau))}{f_Y(F_Y^{-1}(\tau))} d\tau \right).$$

Note that this is the componentwise integral over the Hadamard derivative of the quantile operator from the proof of Proposition 3. Applying Theorem 2, we obtain that

$$\sqrt{n}(S(\hat{\mathbf{F}}) - S(\mathbf{F})) \Rightarrow \mathbb{H}_o.$$

Finally, we have that

$$\sqrt{n}(\mu(\hat{\mathbf{F}}) - \mu(\mathbf{F})) \Rightarrow \mathbb{H}_o(1)$$

because  $\mu(\cdot) = S(\cdot)(1)$ . Taking the last results together, the first statement of the proposition follows from Slutsky's Theorem. The validity of the bootstrap then follows again directly from Theorem 3.  $\square$

## A.9 Proof of Proposition 5

From the continuous mapping theorem, it follows that

$$\sqrt{n}(\hat{\mathbf{G}} - \mathbf{G}) \Rightarrow \int_0^1 \mathbb{L}_o(p) dp.$$

Since  $\mathbb{L}_o$  is a Gaussian process, the term on the right-hand-side is a normally distributed random variable, with mean zero and variance as given in the proposition. The validity of the bootstrap follows again from Theorem 3.  $\square$

## A.10 Proof of Proposition 6

Our proof follows essentially the same lines as the one of Proposition 3 in Barrett and Donald (2003). First, note that the map  $\phi \mapsto \mathcal{D}_j(\cdot, \phi)$  is a linear functional of a Hadamard differentiable mapping. This follows by induction, since  $\mathcal{D}_1(\cdot, \phi)$  is the identity and thus Hadamard differentiable, and the integral operator that transforms  $\mathcal{D}_{j-1}(\cdot, \phi)$  into  $\mathcal{D}_j(\cdot, \phi)$  is linear. It thus follows from Theorem 2 that

$$\sqrt{n}\mathcal{D}_j(\cdot, \hat{F}_Y^* - \hat{F}_Y) = \mathcal{D}_j(\cdot, \sqrt{n}\hat{\Delta}_F) \Rightarrow \mathcal{D}_j(\cdot, (1, -1)\mathbb{F}_o).$$

Furthermore, since the map  $\phi \mapsto \sup_y |\phi(y)|$  is continuous, it follows from the CMT that

$$KS_j \xrightarrow{d} \sup_y \mathcal{D}_j(y, (1, -1)\mathbb{F}_o) = \mathbb{KS}_j.$$

On the other hand, invoking similar arguments and Theorem 3, we obtain for the bootstrapped test statistic that

$$KS_{j,b} \xrightarrow{d} \mathbb{KS}_j, \quad (\text{A.5})$$

conditional on the sample, in probability. Using the same arguments as in Barrett and Donald (2003, p.102), the distribution of  $KS_{j,b}$  is absolutely continuous on  $(0, \infty)$ , and  $c_j(\alpha)$  defined by  $P(KS_j > c_j(\alpha)) = \alpha$  is finite for  $\alpha < 1/2$ . Then the event that  $\hat{p}_j < \alpha$  is equivalent to the event  $KS_j > \hat{c}_j(\alpha)$ , where

$$\inf \{t : P_b(KS_{j,b} > t) > \alpha\} = \hat{c}_j(\alpha) \xrightarrow{P} c_j(\alpha)$$

by (A.5). Then, under the least favorite null hypothesis,

$$\begin{aligned} \lim_{n \rightarrow \infty} P(\text{reject } H_0^j) &= \lim_{n \rightarrow \infty} P(KS_j > \hat{c}_j(\alpha)) \\ &= \lim_{n \rightarrow \infty} P(KS_{j,b} > c_j(\alpha)) + \lim_{n \rightarrow \infty} (P(KS_j > \hat{c}_j(\alpha)) - P(KS_j > c_j(\alpha))) \\ &= \alpha + \lim_{n \rightarrow \infty} P(KS_j \in (\hat{c}_j(\alpha), c_j(\alpha))) \\ &= \alpha \end{aligned}$$

since  $\hat{c}_j(\alpha) \xrightarrow{P} c_j(\alpha)$ . This proves assertion (i) of the proposition. Under the alternative, there is an additional drift term such that the distribution of  $KS_{j,b}$  diverges, and hence  $\lim_{n \rightarrow \infty} P(\text{reject } H_0^j) = 1$  in this case, since  $c_j(\alpha)$  is finite. This proves assertion (ii).  $\square$

## B Influence Function of the Lorenz Curve

The influence function  $\psi^L$  of the bivariate Lorenz process has a lengthy expression, but can be calculated using standard rules of calculus. First, one has to compute the influence function  $\psi^H$  of the Gaussian process  $\mathbb{H}_o$  defined in Proposition 5 by integrating over the influence function  $\psi^Q$  of the quantile process.

It then follows from the product rule that

$$\begin{aligned}\psi_1^L(p, Z_i) &= \frac{1}{\mu_Y^*} \psi_1^H(p, Z_i) - \frac{L_Y^*(p)}{\mu_Y^*} \psi_1^H(1, Z_i), \\ \psi_2^L(p, Z_i) &= \frac{1}{\mu_Y} \psi_2^H(p, Z_i) - \frac{L_Y(p)}{\mu_Y} \psi_2^H(1, Z_i).\end{aligned}$$

Using the definition that  $\alpha(p, x) = Q_Y^*(p)F_{Y|X}(Q_Y^*(p), x) - \mathbb{E}(\mathbb{I}\{Y \leq Q_Y^*(p)\}Y|X = x)$ , the bivariate function  $\psi^H$  can be written as

$$\begin{aligned}\psi_1^H(p, Z_i) &= (pQ_Y^*(p) - H_Y^*(p)) - \alpha(p, X_i^*) \\ &\quad - \frac{f_X^*(X_i)}{f_X(X_i)} (\mathbb{I}\{Y \leq Q_Y^*(p)\}(Q_Y^*(p) - Y_i) - \alpha(p, X_i)), \\ \psi_2^H(p, Z_i) &= (pQ_Y(p) - H_Y(p)) - (\mathbb{I}\{Y \leq Q_Y(p)\}(Q_Y(p) - Y_i)),\end{aligned}$$

where  $H_Y(p) = \int_0^\tau Q_Y(\tau)d\tau$  and  $H_Y^*(p)$  is defined analogously. Noting that  $\alpha(1, x) = E(Y|X = x)$ , we furthermore obtain that

$$\begin{aligned}\psi_1^H(1, Z_i) &= (\bar{y} - \mu_Y^* - \alpha(1, X_i^*)) - \frac{f_X^*(X_i)}{f_X(X_i)} (\bar{y} - Y_i - \alpha(1, X_i)), \\ \psi_2^H(1, Z_i) &= Y_i - \mu_Y,\end{aligned}$$

which completes the description of the components of  $\psi^L$ .

## References

- ABADIE, A. (2002): “Bootstrap Tests for Distributional Treatment Effects in Instrumental Variable Models,” *Journal of the American Statistical Association*, 97(457), 284–293.
- ABREVAYA, J. (2001): “The effects of demographics and maternal behavior on the distribution of birth outcomes,” *Empirical Economics*, 26(1), 247–257.
- AKRITAS, M., AND I. VAN KEILEGOM (2001): “Non-parametric Estimation of the Residual Distribution,” *Scandinavian Journal of Statistics*, 28(3), 549–567.
- ALMOND, D., K. CHAY, AND D. LEE (2005): “The Costs of Low Birth Weight,” *The Quarterly Journal of Economics*, 120(3), 1031–1083.
- ANDERSON, G. (1996): “Nonparametric Tests of Stochastic Dominance in Income Distributions,” *Econometrica*, 64(5), 1183–1193.
- ATKINSON, A. (1970): “On the Measurement of Inequality,” *Journal of Economic Theory*, 2(3), 244–263.
- BARRETT, G., AND S. DONALD (2003): “Consistent Tests for Stochastic Dominance,” *Econometrica*, 71(1), 71–104.
- (2009): “Statistical Inference with Generalized Gini Indices of Inequality, Poverty, and Welfare,” *Journal of Business and Economic Statistics*, 27(1), 1–17.

- BHATTACHARYA, D. (2007): “Inference on inequality from household survey data,” *Journal of Econometrics*, 137(2), 674–707.
- BLACK, S., P. DEVEREUX, AND K. SALVANES (2007): “From the Cradle to the Labor Market? The Effect of Birth Weight on Adult Outcomes,” *The Quarterly Journal of Economics*, 122(1), 409–439.
- BLAU, F., AND L. KAHN (1997): “Swimming Upstream: Trends in the Gender Wage Differential in the 1980s,” *Journal of Labor Economics*, 15(1), 1–42.
- BOSQ, D. (1998): *Nonparametric Statistics for Stochastic Processes: Estimation and Prediction*. Springer.
- CHERNOZHUKOV, V., I. FERNANDEZ-VAL, AND B. MELLY (2009): “Inference on Counterfactual Distributions,” Working Paper.
- CHESHER, A. (2003): “Identification in Nonseparable Models,” *Econometrica*, 71(5), 1405–1441.
- DAVIDSON, R., AND J. DUCLOS (2000): “Statistical Inference for Stochastic Dominance and for the Measurement of Poverty and Inequality,” *Econometrica*, 68(6), 1435–1464.
- DiNARDO, J., N. FORTIN, AND T. LEMIEUX (1996): “Labor Market Institutions and the Distribution of Wages, 1973–1992: A Semiparametric Approach,” *Econometrica*, 64(5), 1001–1044.
- DONALD, S., D. GREEN, AND H. PAARSCH (2000): “Differences in Wage Distributions between Canada and the United States: An Application of a Flexible Estimator of Distribution Functions in the Presence of Covariates,” *Review of Economic Studies*, 67(4), 609–633.
- FAN, J., AND I. GIJBELS (1996): *Local Polynomial Modelling and Its Applications*. Chapman & Hall/CRC.
- FIRPO, S., N. FORTIN, AND T. LEMIEUX (2009): “Unconditional Quantile Regressions,” *Econometrica*, 77(3), 953–973.
- GASSER, T., H. MÜLLER, AND V. MAMMITZSCH (1985): “Kernels for Nonparametric Curve Estimation,” *Journal of the Royal Statistical Society. Series B (Methodological)*, 47(2), 238–252.
- GOSLING, A., S. MACHIN, AND C. MEGHIR (2000): “The Changing Distribution of Male Wages in the UK,” *Review of Economic Studies*, 67(4), 635–666.
- HODERLEIN, S., AND E. MAMMEN (2007): “Identification of Marginal Effects in Nonseparable Models Without Monotonicity,” *Econometrica*, 75(5), 1513–1518.
- HÄRDLE, W., P. JANSEN, AND R. SERFLING (1988): “Strong Uniform Consistency Rates for Estimators of Conditional Functionals,” *The Annals of Statistics*, 16(4), 1428–1449.
- ICHIMURA, H., AND C. TABER (2002): “Semiparametric Reduced-Form Estimation of Tuition Subsidies,” *American Economic Review*, 92(2), 286–292.

- IMBENS, G., AND W. NEWEY (2009): “Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity,” *Econometrica*, forthcoming.
- KOENKER, R., AND K. HALLOCK (2001): “Quantile regression,” *Journal of Economic Perspectives*, 15(4), 143–156.
- LINTON, O., E. MAASOUMI, AND Y. WHANG (2005): “Consistent Testing for Stochastic Dominance under General Sampling Schemes,” *Review of Economic Studies*, 72(3), 735–765.
- MACHADO, J., AND J. MATA (2005): “Counterfactual decomposition of changes in wage distributions using quantile regression,” *Journal of Applied Econometrics*, 20(4), 445–465.
- MATZKIN, R. (2003): “Nonparametric Estimation of Nonadditive Random Functions,” *Econometrica*, 71(5), 1339–1375.
- MCFADDEN, D. (1989): “Testing for Stochastic Dominance,” *Studies in the Economics of Uncertainty (in honor of J. Hadar)*, Springer-Verlag.
- MELLY, B. (2005): “Decomposition of differences in distribution using quantile regression,” *Labour Economics*, 12(4), 577–590.
- NEWEY, W. (1994a): “Kernel Estimation of Partial Means and a General Variance Estimator,” *Econometric Theory*, 10(2), 233–253.
- (1994b): “The Asymptotic Variance of Semiparametric Estimators,” *Econometrica*, 62(6), 1349–1382.
- NOLAN, D., AND D. POLLARD (1987): “U-processes: rates of convergence,” *Annals of Statistics*, 15(2), 780–799.
- PAKES, A., AND D. POLLARD (1989): “Simulation and the Asymptotics of Optimization Estimators,” *Econometrica*, 57(5), 1027–1057.
- SHERMAN, R. (1994): “Maximal Inequalities for Degenerate U-Processes with Applications to Optimization Estimators,” *The Annals of Statistics*, 22(1), 439–459.
- STOCK, J. (1989): “Nonparametric Policy Analysis,” *Journal of the American Statistical Association*, 84(406), 567–575.
- (1991): “Nonparametric Policy Analysis: An Application to Estimating Hazardous Waste Cleanup Benefits,” *Nonparametric and Semiparametric Methods in Econometrics and Statistics: Proceedings of the Fifth International Symposium in Economic Theory and Econometrics*, Cambridge.
- VAN DER VAART, A. (2000): *Asymptotic Statistics*. Cambridge University Press.
- VAN DER VAART, A., AND J. WELLNER (1996): *Weak Convergence and Empirical Processes*. Springer.