Multivariate Distribution Regression

Jonas Meier *

October 3, 2023

First version: 8.9.2020

This paper introduces multivariate distribution regression (MDR), a semi-parametric approach to estimate the joint distribution of outcomes conditional on covariates. The method allows studying complex dependence structures and distributional treatment effects without making strong, parametric assumptions. I show that the MDR coefficient process converges to a Gaussian process and that the bootstrap is consistent for the asymptotic distribution of the estimator. Methodologically, MDR contributes by offering the analysis of many functionals of the multivariate CDF, including counterfactual distributions. Compared to existing models, the contribution of MDR is its flexibility - it requires weak assumptions, is not affected by the curse of dimensionality, and does not require setting tuning parameters. Simulation studies show that MDR's flexibility helps reduce potential biases at moderate costs of increased variances. Finally, I analyze shifts in spousal labor supply in response to a health shock. I find that if low-income individuals receive disability insurance benefits, their spouses respond by increasing their labor supply. The opposite holds for high-income households, likely because they can afford to work fewer hours and look after their partner.

Keywords: Distribution regression; joint distribution; distributional treatment effects

JEL: C14, C21

Acknowledgements: I would like to acknowledge the financial support from the Swiss National Science Foundation for the grant 165621. The instructive comments and valuable suggestions of Blaise Melly, Kaspar Wüthrich, Arturas Juodis, and Lina Zhang were highly appreciated. I would also like to thank participants at the seminars at the University of Bern, University of Fribourg, University of Groningen, at the Econometric Society's European Winter Meeting 2020, the Alumni Conference in Gerzensee and the DAGStat Conference 2022 for their inputs. Explicitly, the remarks of Tamara Bischof, Martina Pons, Severin Lenhard, Dino Collati, Armando Näf, Nadia Ceschi and Marc Brunner have been highly profitable. Further, I am grateful for having the possibility to use Alexis Derumigny's R Codes for conditional copulas. Finally, I am thankful to have been granted access to the WiSIeR data, a collection of datasets provided by the Swiss Federal Social Insurance Office.

*Amsterdam School of Economics, University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands. Email: j.c.meier@uva.nl.

1 Introduction

Researchers often aim to estimate the effect of covariates on the joint distribution of outcomes (see Patton, 2012, for an overview). Such situations arise in settings where observables affect the dependence between outcomes. For instance, this is the case for the division of labor supply within households. The rising correlation between spouses' incomes depends on the allocation of housework and is an essential driver of inequality (Hyslop, 2001; Schwartz, 2010). The allocation of labor, in turn, is a function of parenthood, bargaining power, and norms (e.g. Kleven et al., 2019; Kühhirt, 2011). Another example is equity price dynamics in multiple markets. The co-movement of asset returns depends on a list of factors such as the degree of financial integration, market cycles, or country-specific characteristics (e.g. Christoffersen et al., 2012; Aloui et al., 2011). In both cases, explicitly allowing the dependence structure to vary with the regressors may provide additional insights that univariate approaches would miss. In this spirit, this paper derives a flexible estimator of the multivariate cumulative distribution function (CDF) conditional on a set of covariates.

Multivariate Distribution Regression (MDR) builds on the theory of univariate Distribution Regression (DR) initially introduced by Williams and Grizzle (1972) and applied to ordered outcomes by Jung (1996). Foresi and Peracchi (1995) were the first to establish pointwise convergence at a finite number of thresholds. Building on these results, Chernozhukov et al. (2013) proved that a functional central limit theorem holds for univariate conditional CDFs estimated using DR. This paper departs from their setting and extends univariate DR to the multivariate case. While the results in Chernozhukov et al. (2013) are based on theorems for approximate Z-estimators, I make use of exact Z-estimators, as quantile regression methods play no part in this analysis. In its simplest form and analog to univariate DR, MDR reduces to the empirical multivariate CDF when using only a constant as a regressor. Thus, MDR generalizes two well-studied estimators: (i) the empirical multivariate CDF by including covariates and (ii) univariate DR by considering multiple outcomes. In contrast to univariate DR, MDR provides information on the dependence structure, for instance, the correlation matrix. Considering the general case, the theoretical contribution of this paper is threefold. I show (i) that the MDR coefficient process converges to a mean-zero Gaussian process, (ii) that the corresponding variance is consistently estimated by a bootstrap technique, and (iii) that functionals of the fitted CDF are consistently estimated by the functional delta method. The last result is of great relevance from a methodological point of view. Essentially, MDR is attractive as it offers many novel possibilities to analyze the joint CDF without strongly restricting the latter parametrically. In this regard, the main contribution of the present paper is its relevance for applied research, which is underlined by the examples in section 2 and 5.

At the model's core, MDR estimates the joint CDF of multiple outcomes conditional on covariates. Considering a bivariate case, denote the outcomes by $Y_{1,i}$ and $Y_{2,i}$ for observation i. The joint CDF is the probability of $Y_{1,i}$ and $Y_{2,i}$ being smaller than some specified values, say t_1 and t_2. Formally, this is $F_{Y_i}(t) = P(Y_{1,i} \leq t_1, Y_{2,i} \leq t_2)$, where t is the vector of thresholds. Equivalently, $F_{Y_i}(t)$ is the expected value of the binary variable $E[\mathbf{1}(Y_{1,i} \leq t_1, Y_{2,i} \leq t_2)]$. Most applied research is interested in how regressors affect this probability. A natural possibility to model $E[\mathbf{1}(Y_{1,i} \leq t_1, Y_{2,i} \leq t_2)|X_i]$ is to use a binary regression such as a logistic or probit model. In the following, denote the value of the conditional CDF at t_1 and t_2 by $F_{Y_i|X_i}(t|X_i)$. Essentially, the MDR estimator models this conditional expectation at many thresholds. The resulting coefficients on X_i, $\beta(t)$, are allowed to vary with t, which ensures a high degree of flexibility. This modeling approach entails several attractive features. First, the estimator is trivial to implement. Second, each individual regression is tractable and offers various forms of well-known results, such as marginal effects. Third, the obtained fit of the CDF can be used to estimate any feature of the joint distribution. Beyond many univariate statistics, this includes, for instance, quantile functions (QF) at specific locations within the joint distribution, measures of tail dependencies, or transition matrices. Finally, having obtained the full conditional distribution, it is straightforward to compute counterfactual distributions by (i) either integrating over a modified distribution of covariates or (ii) changing the conditional distribution itself. This offers a large variety of potential functionals that can be tailored to the specific case at hand.

I illustrate the model in an application to the division of labor supply within Swiss households. More precisely, I estimate how household labor income changes after the main earner newly receive Disability Insurance Benefits (DIB). Generally, receiving DIB is associated with a lower labor supply (e.g. Autor et al., 2016). Due to fiscal debates on who should be subsidized, this effect has drawn much attention in recent years.[1] Motivated by the study of Autor et al. (2019), I expect that spouses expand their labor supply and partly compensate for the loss of household income which the insurance might not fully cover. MDR offers the possibility to analyze these shifts separately at all parts of the bivariate distribution of labor earnings. I find that the spouses' response crucially depends on their partner's income. If the shock hits low-income main earners, the spousal labor supply increases to compensate for the financial loss of the household. In contrast, spouses of high-income individuals reduce their labor supply. Likely, these individuals

[1]For instance, changing the rules on eligibility of DIB was a much-disputed topic in Britain (Walker and Elgot, 2017). Recently, in relation to Covid-19, Canadian politicians disagreed over emergency DIB (Canadian Press, 2020).

take care of their partners and the household chores as they can afford to work fewer hours (as suggested by Lee, 2020).

In recent years, many studies have considered extensions to univariate DR as introduced in Chernozhukov et al. (2013). For instance, this includes introducing two-way fixed effects into DR models (Chernozhukov et al., 2020a) or estimating structural functions in nonseparable triangular models (Chernozhukov et al., 2020b). Both studies highlight the flexibility of the underlying framework, which is of great importance for applied research as fixed effects and endogenous regressors are prevalent in many settings. Similar extensions seem feasible in the context of MDR, yet, the present paper leaves this for future research. Ćevid et al. (2020) have suggested a fully nonparametric approach to multivariate DR using random forests. Like similar nonparametric methods discussed below, this estimator might suffer from the curse of dimensionality. Most recently, Wang et al. (2022) proposed to estimate bivariate distributions using a factorization approach based on two univariate DRs. However, the authors impose that one outcome variable (Z) is discrete and linearly shifts the other outcome (Y). This is restrictive, as, for instance, no interaction between the first outcome and the covariates is present. Thus, it is by construction impossible that Y and Z are positively correlated for some X and negatively for others. Further, note that the conditioning outcome variable needs to have the same effect if switched from 0 to 1 as from 2 to 3. Similar restrictions are not present with MDR. The comparison between the approaches below shows that these restrictions might be crucial.

As an alternative to MDR, researchers may consider quantile regression (QR) as a tool to model the conditional distribution of an outcome. Since the seminal paper by Koenker and Bassett (1978), QR has developed into a standard method to analyze heterogeneity. Several authors extended the principles to the multivariate case. However, this is a nontrivial exercise due to the lack of canonical ordering in higher dimensions. In search of a unified approach, Zuo and Serfling (2000) introduced desirable properties of statistical depth functions. Recently, Chernozhukov et al. (2017) extended the QR framework to the multivariate setting and provided results on multivariate quantiles. Compared to QR, DR naturally generalizes to multivariate tasks and neatly handles mixed or discrete outcomes (Chernozhukov et al., 2019). For instance, wages or labor market participation are typically non-continuously distributed. Thus, researchers may favor DR in such applications.

More broadly, the derived estimator fits in a literature concerned with the convergence of empirical processes. Dudley (1966) has been the first to derive a theory on multivariate empirical distributions, and numerous authors extended this result to more involved settings (e.g. Delattre and Roquain, 2016). In principle, parametric and nonparametric approaches might be suitable for estimating a multivariate CDF. For instance, Gijbels et al. (2011) proposed nonparametric estimates of a copula model. However, as the number of regressors increases, these models become infeasible in practice because they suffer from the curse of dimensionality (Fermanian and Lopez, 2018). Similarly, the nonparametric approach of Bouzebda and Nemouchi (2019) using U-processes is likely to suffer from the same drawback. One might resolve the issue by imposing a structural form of the distribution. In this spirit, recent studies aiming to model conditional CDFs tackled the issue using copula models (Fermanian and Lopez, 2018; Portier and Segers, 2018). In general, copulas are attractive in this setting due to the possibility to separately specify the marginals and the dependence structure, the copula itself, ensuring a high degree of flexibility (e.g. Patton, 2012). For instance, Klein et al. (2019) proposed a setting where the estimation of the conditional CDF is replaced by the estimation of a monotonically increasing transformation function. This simplifies the estimation and the derived inference theory. Still, the choice of the transformation function remains specific to the case at hand. Compared to these competitors, MDR is set in between the nonparametric and parametric methods. Although the assumption on the link function, the analog to the transformation function of Klein et al. (2019), is parametric, the regressors can flexibly affect the CDF throughout the distribution. Thus, using MDR, it is feasible to account for many confounders and reduce the risk of dimensionality issues. Further, the practical advantages of MDR are twofold. First, the conditional, multivariate CDF implies all copula-type parameters of the distribution, while the opposite does not hold. For an overview of potential parameters of interest, see Callaway et al. (2021). Second, the risk of misspecification is lower compared to copula models as the assumption on the copula is crucial and may be too restrictive (Zimmer, 2012; Ho et al., 2015). In a simplistic simulation setting, I show that MDR performs equally well as copula models, even in cases with only one regressor.

The remainder of this paper is structured as follows. Section 2 introduces the model and presents several examples illustrating the advantages of MDR and counterfactual CDFs. Section 3 derives the asymptotic theory of the estimator and establishes further relevant results. A simulation example is set up in section 4. Section 5 presents the application, and section 6 concludes.

2 Model

This section introduces the approach of MDR by first presenting how the multivariate CDF is modeled and then focusing on desired functionals of the CDF, primarily counterfactual distributions. To illustrate the wide range of potential applications, I present several use cases of MDR in the second half of this section. In the following, let $F_{Y_i|X_i}(t|X_i)$ be the multivariate CDF of the d-dimensional response vector $Y_i = (Y_{1,i}, .., Y_{d,i})$, where $t \in \mathcal{T} = \mathbb{R}^d$ is a vector of thresholds, X_i is a set of K regressors and $i = 1, ..., N$ indexes the observations. Throughout, the multivariate CDF is modeled by the assumption of a parametric link function Λ and a coefficient vector $\beta(t)$, i.e.

$$F_{Y_i|X_i}(t|X_i) = \Lambda(X_i'\beta(t)). \tag{1}$$

Typically, Λ is either a complementary log-log, logistic, or probit function. At a given vector of thresholds, this model is no different from running a standard binary regression of $\mathbf{1}(Y_i \leq t)$ on the regressors X_i. To see this, note that by definition, a conditional CDF can be written as $F_{Y_i|X_i}(t|X_i) = P(Y_i \leq t|X_i) = E[\mathbf{1}(Y_i \leq t)|X_i]$. Consequentially, at each t, one obtains a set of estimated coefficients $\hat{\beta}(t)$. In a nutshell, MDR is concerned with the uniform convergence of all these estimates, i.e., the limiting process of the estimated coefficients. Intuitively, one can think of estimating the model in equation (1) at infinitely many thresholds and treating the obtained coefficients as a function of t.

The core of this modeling approach is the link function Λ. In the case of univariate DR, it can be shown that using a complementary log-log link function, the model nests the duration model by Cox (1972). An analog location-shift representation is not available in the case of MDR, yet, the approach entails other attractive features. For instance, we can derive the following relation by imposing a logistic link function and assuming independence of $Y_{1,i}$ and $Y_{2,i}$.

$$\begin{aligned} F_{Y_{1,i},Y_{2,i}|X_i}(t|X_i) &= \Lambda(X_i'\beta(t)) \\ &= \Lambda(X_i'\beta_1(t_1))\Lambda(X_i'\beta_2(t_2)), \end{aligned}$$

where $\beta_1(t_1)$ and $\beta_2(t_2)$ are the coefficient vectors of both univariate DRs. In this particular case, it holds that $X_i'\beta(t) = -\ln\left[\exp(-X_i'\beta_1(t_1)) + \exp(-X_i'\beta_2(t_2)) + \exp(-X_i'(\beta_1(t_1) + \beta_2(t_2))\right]$, i.e. $\beta(t)$ has a representation based on univariate DR coefficients only. A similar result is available for the complementary log-log link function.

The underlying implications of this modeling approach, including the assumptions on Λ, are discussed in detail in section 3.1. While the estimated coefficient vector is interesting on its own, in many cases, the focus will lie on the implied distribution. In particular, it is feasible to obtain the average value of the multivariate CDF by integrating over the *observed* covariate distribution $F_{X_i}(X)$. Formally, this is equivalent to

$$F_{Y\langle\cdot|\cdot\rangle}(t) = \int_{\mathcal{X}_i} F_{Y_i|X_i}(t|X)dF_{X_i}(X), \tag{2}$$

where the subscript $\langle\cdot|\cdot\rangle$ denotes that neither the conditional distribution $F_{Y_i|X_i}(t|X)$ nor the covariate distribution $F_{X_i}(X)$ has been changed. Note that the resulting distribution no longer depends on X_i. Building on equation (2), one could integrate over a modified covariate distribution to obtain a counterfactual CDF (see section 3.3 for all types of counterfactual distributions). To illustrate the value of this artificial distribution, suppose we are interested in the effect of a binary treatment D_i that is part of X_i. A natural choice of two counterfactuals would be the CDF in the presence or absence of D_i, i.e. $D_i = 0$ or $D_i = 1$. Intuitively, this answers "How would the CDF look like if the effect of D_i treated everyone (no one)?". Denoting the covariate distribution with all values of D_i set to $d \in \{0, 1\}$ by $F_{X_{i,D_i=d}}(X)$, the counterfactual CDFs and the resulting, distributional treatment effects can be written as

$$F_{Y\langle\cdot|D=d\rangle}(t) = \int_{\mathcal{X}_{i,D_i=d}} F_{Y_i|X_i}(t|X)dF_{X_{i,D_i=d}}(X) \text{ and} \tag{3}$$

$$\Delta(t) = g(F_{Y\langle\cdot|D=1\rangle}(t)) - g(F_{Y\langle\cdot|D=0\rangle}(t)), \tag{4}$$

where the choice $g(\cdot)$ depends on the research question. Below, I propose to consider the CDF of one outcome at specific locations of the joint distribution. Other choices of $g(\cdot)$ include tail dependence measures, transition, or variance-covariance matrices. Naturally, $\Delta(t)$ implies univariate, distributional effects and aggregate effects too, for instance, the means and variances of Y_i. Note that the comparison of the two CDFs has a causal interpretation when a standard conditional independence assumption (CIA) is imposed, i.e., that the treatment is randomly assigned once conditioned on the covariates.[2] To fix ideas and highlight the added value of MDR, I present several examples in the following. First, I demonstrate the flexibility of MDR imposing a known outcome distribution. Second, I compare MDR to competitive copula approaches using real-world data on stock returns. Finally, I sketch two additional examples in applied research where MDR could benefit.

Example 1 : Clayton Copula Suppose the outcome distribution is generated by a Clayton copula and marginals, which follow standard normal distributions. Formally,

[2]For a detailed discussion on when counterfactuals do have causal interpretation, see Chernozhukov et al. (2013, section 2.3).

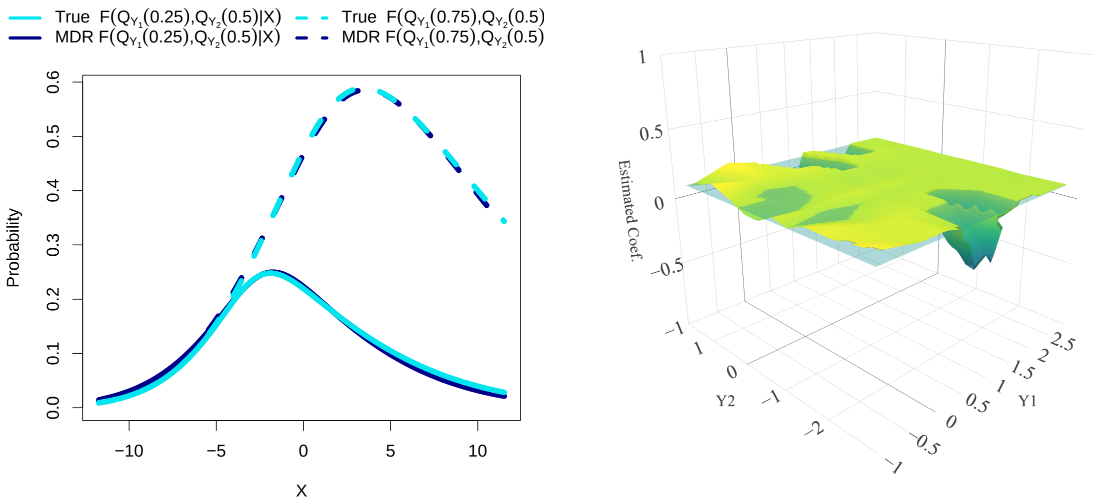
this is

$$F_{Y_{1,i},Y_{2,i}|X_i}(y_1, y_2|x) = C\left(F_{Y_{1,i}|X_i}(y_1|x), F_{Y_{2,i}|X_i}(y_2|x)\right)$$
$$= \left(\Phi(y_1|x)^{-\theta} + \Phi(y_2|x)^{-\theta} - 1\right)^{-\frac{1}{\theta}}, \tag{5}$$

where $\Phi(\cdot)$ is the standard normal distribution, $X_i \sim N(0, 5)$ and, for simplicity, $\theta = 3$ is assumed to be independent of X_i. Further, assume that $Y_{j,i} = \beta_{0,j} + \beta_{1,j}X_i + \epsilon_j$ for $j = 1, 2$ where both ϵ_j follow a standard normal distribution and ensure that joint distribution of $Y_{1,i}$ and $Y_{2,i}$ follows (5). This particular case is tractable as it implies a known link function and is simple enough to demonstrate the generality of the MDR estimator. Next, I consider how well MDR is able to estimate the underlying joint distribution. For this purpose, I employ a maximum likelihood procedure where θ is estimated in the first step based on the least squares residuals from $Y_{1,i}$ and $Y_{2,i}$ on X_i, respectively. In a second step, $\beta_{0,1}, \beta_{1,1}, \beta_{0,2}$ and $\beta_{1,2}$ are estimated based on the likelihood function given θ. Panel (a) of Figure 1 displays the value of the CDF at two sets of thresholds conditional on X_i. We observe that across all values of X_i, the MDR estimator matches the underlying distribution from (5) closely. This implies that the correctly specified MDR estimator nests the Clayton copula in (5) as a special case. To see this, note that all four coefficients were assumed to be constant across the distribution. However, the MDR estimator allows for the coefficients to depend on y_1 and y_2, i.e., the location of the distribution. This is visualized in panel (b) of Figure 1 where all estimated values of $\beta_{1,1}$ are displayed. The estimates narrowly float around the true value of .1, which is represented by the transparent layer. At each gridpoint, the MDR estimator would allow for a different $\beta_{1,1}$. Thus, all estimates together form a basis to test whether the parameters are, in fact, constant. In this regard, MDR nests the distribution under (5) with constant parameters as a strict special case. Note that while section A.3 sketches testing possibilities, developing a test on the multivariate grid is beyond the scope of the present paper.

Example 2 : SP500 & FTSE100 The interdependence of stock prices in multiple markets has long been recognized (e.g. Christoffersen et al., 2012; Aloui et al., 2011; Chavas, 2020). In short, positive correlations arise across countries and markets. These correlations depend on numerous factors, such as market cycles, the type of goods, and the economy's stage of development. The stylized example below aims at two goals: (i) estimating the mutual risk of losses and (ii) demonstrating the MDR's flexibility. The former can be conceptualized by the tail dependence implied by the joint distribution of negative returns. Taking two well-studied indices, the SP500 and the FTSE100, denote

Figure 1: Correctly specified MDR



(a) Estimated CDF (b) Estimates of $\beta_{1,1}$

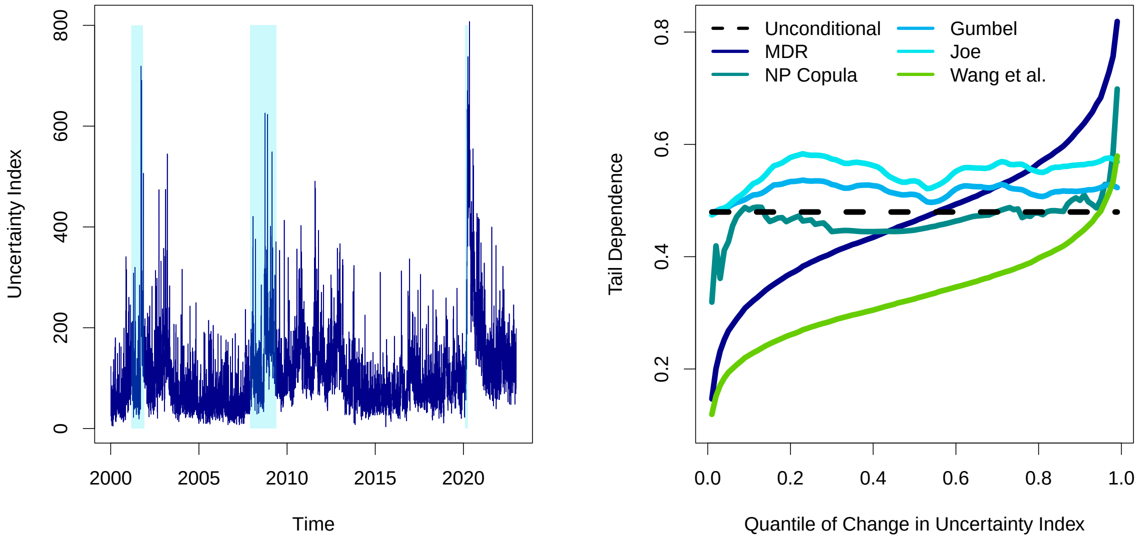
Notes: This figure is based on 5000 observations drawn from (5) with $Y_{1,i} = 1 + .1X_i + \epsilon_1$, $Y_{2,i} = -.5 - .2X_i + \epsilon_2$ and $\theta = 3$. For panel (b), a grid of 10×10 (quantile values from .05 to .95) was used.

the tail dependence by

$$\begin{aligned}\bar{\chi} &= \frac{2\ln\left(\Pr(nr_{US} > Q_{95,US}\right)}{\ln\left(\Pr(nr_{US} > Q_{95,US},\ nr_{UK} > Q_{95,UK})\right)} - 1 \\ &= \frac{2\ln(.05)}{\ln\left(1 + F_{US,UK}(Q_{95,US}, Q_{95,UK}) - F_{US,UK}(\infty, Q_{95,UK}) - F_{US,UK}(Q_{95,US}, \infty)\right)} - 1,\end{aligned} \tag{6}$$

where nr_{US} and nr_{US} indicate the negative returns, $Q_{95,US}$ and $Q_{95,UK}$ are their 95th quantiles and $F_{US,UK}(us, uk)$ represents their joint distribution. Note that the tail dependence on the 95% level introduced in equation (6) is a direct function of the joint, i.e., $\bar{\chi} = g(F_{US,UK}(us, uk))$. It is straightforward to allow the CDF in (6) to depend on covariates. Using a single regressor, log changes in the FRED economic uncertainty index Δunc, we can model the tail dependence as a function of the covariate. Specifically, we generate a counterfactual distribution for each percentile of Δunc according to equation (3). We expect substantial increases in uncertainty to lead to (i) higher losses and (ii) an increased tail dependence, as global uncertainty is likely to affect the markets jointly. Having filtered all indices using an ARMA(1,1) each, I run five different models: The MDR, a non-parametric copula model, two parametric models imposing a Gumbel and Joe copula, respectively, and the factorization estimator by Wang et al. (2022). For the latter, note that we must discretize one return series; 25 values were chosen. The copula estimators are introduced in Derumigny and Fermanian (2017). Two issues might affect the fit quality in the present case: (i) potential misspecification because of the parametric restrictions and (ii) the underlying bias-variance trade-off. In particular, the copula models benefit from a reduced variance due to their distributional assumption. However, the MDR estimator can reduce the bias due to its flexibility. The simulation study in section 4 supports these claims.

Figure 2: SP500 & FTSE100



(a) Uncertainty Index (b) $\bar{\chi}$ as a function of uncertainty

Notes: Panel (a) plots the uncertainty index with three major crisis highlighted: The 2001 recession, the 2008 financial crisis and the 2020 covid pandemic. In panel (b), the tail dependence is modelled as a function of changes in the uncertainty index. For the non-parametric copula model, a bandwidth of .5 was chosen. For the factorization estimator by Wang et al. (2022), 7 gridpoints between .95 and 1 were used. Confidence bands are omitted to improve readability.

The results in Figure 2 are based on daily data from January 2000 to December 2022. From panel (b), we find that in moderate times the tail dependence is close to 0.4, which implies a very weak relation of the indices. On the other hand, when the uncertainty is high, we observe a substantial tail dependence, considerably higher than estimated previously (Poon et al., 2004). This is in line with our expectations stated above. Comparing the results across models, we note that all estimates follow a similar pattern. The parametric copula models are less sensitive to extreme values of uncertainty. To some extent, this is because of their comparably stronger assumptions; partly, this could be due to misspecification (Zimmer, 2012). The non-parametric copula and the MDR entail more flexibility; both pick up a more variable pattern. However, the chosen bandwidth for the former restricts this flexibility. Finally, note that the estimates for the model by Wang et al. (2022) depend on two practical issues: The grid's thinness and the data's discreteness can introduce a bias. The latter becomes apparent as the unconditional distribution does not imply the average tail dependence. Also, the fact that the second outcome might only shift the first linearly at one gridpoint restricts the dependence structure. This does not seem to matter in the present case but can be seen in more complicated settings, as in section 5. In contrast to example 1, these employed models are generally not nested. Therefore, we cannot directly test the validity of the parametric assumptions. However, the MDR is most likely to provide an accurate estimate as it is the most flexible approach, thus reducing the risk of a bias. In summary, MDR is attractive because it does not require a stand on comparably strong parametric assumptions or tuning parameters and does not suffer from the curse of dimensionality.

Example 3: **Bivariate Labor Supply** Family labor supply has gained much attention due to its relevance for intra-family and aggregate inequality (Eika et al., 2019; Hyslop, 2001; Schwartz, 2010). Consider two spouses who both participate in the labor market. Let $Y_{1,i}$ and $Y_{2,i}$ denote the labor income of the wives and husbands, respectively. Notably, labor income depends on several factors such as gender, age, education, parenthood, origin, or industry (see, e.g. Blau and Winkler, 2018, for an overview). Thus, one should include these characteristics in a model of $Y_{1,i}$ and $Y_{2,i}$ to avoid misspecification - a task for which MDR is well suited. Within this set of regressors, let us reconsider the effect of a binary variable D_i. For instance, tax reform, winning the lottery, unemployment, or a health shock may serve as treatment variables. Cesarini et al. (2017) find that winning the lottery modestly reduces the labor supply for winners. The reduction is smaller for spouses, which is inconsistent with unitary household models. Using MDR, modeling the multivariate labor supply response could reveal that the impacts depend on the initial earnings of both partners. Thus, while the average effects are small, individuals at the bottom of the distribution may experience considerable changes. To see this, let us consider the effect on $Y_{1,i}$ at specific levels of $Y_{2,i}$, i.e. $Q_j(Y_{2,i}) \leq Y_{2,i} < Q_k(Y_{2,i})$ where $Q_k(Y_{2,i})$ denotes the k^{th} quantile of $Y_{2,i}$. Accordingly, we can obtain $F_{Y_1\langle \cdot | Q_j(Y_{2,i}) \leq Y_{2,i} < Q_k(Y_{2,i}), D\rangle}(t)$ by setting

$$g(F_{Y\langle \cdot | j,k;D\rangle}(t)) = \frac{F_{Y\langle \cdot | D\rangle}(Y_1, Q_k(Y_2))(t) - F_{Y\langle \cdot | D\rangle}(Y_1, Q_j(Y_2))(t)}{F_{Y\langle \cdot | D\rangle}(\infty, Q_k(Y_2))(t) - F_{Y\langle \cdot | D\rangle}(\infty, Q_j(Y_2))(t)}. \tag{7}$$

The definition of the treatment effect directly follows from (4), and the effect on $Y_{2,i}$ can be looked at analogously. In the context of tax reforms, it might be interesting to see how households would react to a change from joint to individual taxation for married couples.

[3] Considering 17 European countries and the US, Bick and Fuchs-Schündeln (2017) show that married women would increase their hours worked by 10% if they were taxed individually. Drawing on these implications, the MDR estimator could answer whether the potential increase in female labor supply differs across the spouses' distribution of earnings. Other applications could focus on the effect of a health shock (Fadlon and Nielsen, 2021), extending unemployment benefits (Nekoei and Weber, 2017), job loss (Halla et al., 2020), worktime regulations (Goux et al., 2014), or retirement benefits (Manoli and Weber, 2016).

Example 4 : Intergenerational Mobility Many studies find that children's economic status depends on their parent's characteristics. How strongly the latter determine the children's education, earnings, and success is usually described as intergenerational mobility. Neighbourhoods (Chetty and Hendren, 2018a,b), migration (Ward, 2020), college allocation (Chetty et al., 2020), and education (Lavy et al., 2022) are among the many factors affecting mobility. In the context of MDR, one could model the childrens' and parents' variables of interest simultaneously. Typically, this literature conceptualizes mobility by looking at transition matrices or upward mobility, i.e., the childrens' probability of attaining a higher value of the outcome than their parents. As shown on page 9 in Callaway et al. (2021) both are a function of the marginal distributions which, in turn, are implied by the joint distribution. Thus, the choice of $g(\cdot)$ naturally follows from the equations in Callaway et al. (2021). Making use of counterfactual distributions analog to equation (3), we could model transition matrices for individuals from poor (rich) neighborhoods, bad (good) colleges, or for different countries of origin. In this regard, the presented framework may deepen existing findings by detecting new sources of heterogeneity.

3 Asymptotic Theory

3.1 Assumptions

This section introduces and discusses the assumptions on the underlying data and the model.

Assumption 1 : Data *The data Y_i, X_i is i.i.d.*

Assumption 2 : Model *The multivariate CDF is modeled by a parametric link function Λ, i.e.*

[3]This is discussed in Switzerland, where married couples pay up to 10% more in taxes than their single counterparts (Peters, 2014). Due to this inequality, there have been numerous attempts to reform the tax system (e.g. Schöchli, 2019). However, the political parties have not reached an agreement (CH Media, 2020).

$$F_{Y_i|X_i}(t|X_i) = \Lambda(X_i'\beta(t)), \tag{8}$$

where $\beta(t)$ is a $K \times 1$ function-valued coefficient vector and X_i is a $K \times 1$ matrix of regressors. The link function Λ is assumed to be either a linear, probit, logistic, complementary log-log or cauchit function. Let $\beta_0(t)$ denote the true parameters.

Assumption 1 is standard for DR models. Note that Assumption 1 can be relaxed as consistency and convergence of Z-estimators are more general (e.g. Kosorok, 2008, p. 246). The model introduced in assumption 2 is semi-parametric in the sense that it requires a parametric link function while allowing the coefficients to vary flexibly with the thresholds. The choice of the link function should be viewed in the light of two arguments. First, in the absence of covariates, the link function does not affect the results and the model generalizes to the non-parametric estimate of the empirical distribution function. Second, if X_i is rich enough, the CDF is approximated arbitrarily well and the parametric form of the link function is irrelevant too.[4]

Additional remarks should be made with respect to assumption 2. First, the estimator for $\beta_0(t)$ defined in equation (8) is a Z-estimator and can be interpreted as a pseudo maximum likelihood estimator (MLE) because it provides the best approximation of the CDF given a specific link function. Second, the model in equation (8) may be misspecified, yet consistent results can still be obtained under mild regularity conditions (see Theorem 2 in Wald, 1949; White, 1982) as the estimator sets in a pseudo-MLE framework. Third, the model allows for the inclusion of a wide range of functionals of X_i, which ensure that the CDF is approximated well. Finally, the distribution function in equation (8) does not need to be continuous as the model also captures mixed or discrete outcome variables.[5] Yet, what is needed to obtain uniform results for the MDR estimator below is that the map $t \mapsto F_{Y_i|X_i}(t|X_i)$ is uniformly continuous as stated in the following assumption 3. Essentially, this assumption ensures that the distance between the grid points on t is indefinitely small, thus establishing a continuum on t. Note that this assumption is analog to condition D (b) in Chernozhukov et al. (2013) and does not require the outcome variables to be continuous as the approximation of the CDF is made pointwise for each

[4] Following the argument in Chernozhukov et al. (2013, p. 2217), let $P_k(X)$ consist of the first p components of a basis in $L^2(\mathcal{X}, P_k)$. Assume that $\Lambda^{-1}\left[F_{Y_i|X_i}(t|X)\right] \in L^2(\mathcal{X}, P_k)$ and $\lambda(z) = \frac{\partial \Lambda(z)}{\partial z}$ is bounded above by $\bar{\lambda}$. Define the squared misspecification error $\delta_p = E\left[\Lambda^{-1}(F_{Y_i|X_i}(t|X)) - P_k(X)'\beta(t)\right] E\left[\Lambda^{-1}(F_{Y_i|X_i}(t|X)) - P_k(X)'\beta(t)\right]'$. Then, it can be shown that $\delta_p \to 0$ as p grows. Thus, $E\left[F_{Y_i|X_i}(t|X) - \Lambda(P_k(X)'\beta(t))\right] E\left[F_{Y_i|X_i}(t|X) - \Lambda(P_k(X)'\beta(t))\right]' \leq \bar{\lambda}\delta_p \to 0$ by weak concavity of Λ.

[5] To see this, suppose that the true CDF $F_{Y_i|X_i}(t|X_i)$ is discrete or mixed. As the approximation is made pointwise at the thresholds t, discrete or mixed outcomes can easily be handled by the binary regressions. In this case, estimating the model in equation (8) reveals several times the same coefficient vector. This is equivalent to estimating the model at a finite number of thresholds. Thus, the coefficient vector is consistently estimated and jointly normally distributed with known variance.

t.

Assumption 3 : Grid *Let $h(f,\tilde{f}) = [\int (f-\tilde{f})^2 dF_{X_i}]^{1/2}$ be a metric on $\mathcal{F}$, where $\mathcal{F}$ is a class of measurable functions including the CDF from assumption 2 as well as the indicators of all the rectangles in $\mathbb{R}^{d_K}$. As a result, $\mathcal{F}$ is totally bounded under h. Assume that the map $t \mapsto F_{Y_i|X_i}(t|X_i)$ is uniformly continuous with respect to the metric h.*

The objective function is defined as an approximate zero of a function $\Psi(\beta,t)$ between two normed spaces. More precisely, $\Psi(\beta,t) : \Theta \times \mathcal{M} \mapsto \Theta$ where $\mathcal{M}$ is an open set containing $\mathcal{T}$ and $\Theta = \mathbb{R}^{d_K}$ is the parameter space which contains β with d_K being the dimension of the regressor matrix. I define the corresponding norms to be Euclidean norm $||\cdot||$ and the infinity norm $||\cdot||_\infty$. By Assumption 1 and since the binary regressions are based on MLE, $\Psi(\beta,t) = P\psi_{\beta,t}$, where P is the probability measure and $\psi_{\beta,t}$ is equivalent to the derivative of the log likelihood, that is

$$\psi_{\beta,t} = \left(\Lambda\left[X_i'\beta(t)\right] - y_i(t)\right)\left(\frac{\lambda\left[X_i'\beta(t)\right]X_i}{\Lambda\left[X_i'\beta(t)\right]\left(1-\Lambda\left[X_i'\beta(t)\right]\right)}\right), \tag{9}$$

where $\lambda(\cdot)$ denotes the derivative of $\Lambda(\cdot)$ and $y_i(t) = \mathbf{1}(Y_{1,i} \leq t_1, ..., Y_{d,i} \leq t_d)$. Accordingly, let $\Psi_n(\beta,t) = \mathbb{P}_n\psi_{\beta,t}$ be the sample estimator, where $\mathbb{P}_n$ is the empirical measure. Thus, the MDR estimator $\hat{\beta}_n$ is the solution to the estimating equations $\Psi_n(\beta,t) = 0$, i.e. satisfying $||\Psi_n(\hat{\beta}_n,t)||_\infty \xrightarrow{p} 0$. The subsequent assumption is concerned with the properties of $\Psi(\beta,t)$ itself. Then, assumption 5 introduces the requirements for the bootstrap to be a valid tool to do inference.

Assumption 4 : Identifiability *At the true values $\beta_0 \in \Theta$, $\Psi(\beta_0,t) = 0$. Further, assume that both, $\Psi(\beta_0,t)$ and $\Psi(\beta,t) : \mathcal{M} \times \Theta \mapsto \Theta$ are one-to-one maps.*

Assumption 5 : Bootstrap *Let $\hat{\beta}_n$ be the MDR estimator and $\hat{\beta}_n^\circ$ be a minimizer of $\sup_{t\in\mathcal{T}}|\Psi_n^\circ(\beta,t)|$ where $\Psi_n^\circ(\beta,t) = \mathbb{P}_n^\circ\psi_{\beta,t}$, and $\mathbb{P}_n^\circ f = n^{-1}\sum_{i=1}^n \frac{\xi_i}{\bar{\xi}} f(X_i)$ denotes the non-parametric or multiplier bootstrap where $\bar{\xi} = n^{-1}\sum_{i=1}^n \xi_i$. Assume that $\xi_1, ..., \xi_n$ are i.i.d positive weights with $0 < \mu = E\xi_1 < \infty$. In the case of the multiplier bootstrap, additionally assume that $0 < \tau = var(\xi_1) < \infty$ and $||\xi_1||_{2,1} < \infty$.*

Assumption 4 requires the objective function to be zero only at the true values $\beta_0(t)$. Thus, assumption 4 ensures that the true parameters are identified. Note that the condition on the maps being one-to-one is stronger than what is needed (Kosorok, 2008, p. 244), however, the assumption simplifies the proofs. In most cases, assumption 4 is seen as a technical requirement which is likely to be met.

With respect to assumption 5, two comments should be made. First, theoretically, the only requirement is that the bootstrapped estimator is an approximate zero of the bootstrapped estimating equation. This allows for numerous forms of bootstraps. The multiplier bootstrap and non-parametric bootstrap with multinomial weights are shown to provide valid results as they capture two important cases in practice. Second, as $\Psi(\cdot)$ can be shown to be strong Glivenko-Cantelli, other designs such as the exchangeable bootstrap would be applicable too.

Before addressing uniform convergence of MDR, note that pointwise convergence is established from standard asymptotic arguments in the framework of maximum likeihood estimation. As mentioned above, the model in equation (8) is estimated as single regressions at every entry of t. For a given t, $y_i(t)$ is regressed on X_i, for instance using a probit model. Thus, at a finite number of t, MLE ensures that $\beta_0(t)$ is consistently estimated. Further, the estimator of $\beta_0(t)$ is $\sqrt{n}$-consistent and asymptotically normally distributed with a known expression for the variance which only depends on the chosen link function. Finally, note that a finite number of MDR estimators at a finite number of thresholds are jointly normally distributed. Yet, in contrast to Foresi and Peracchi (1995) and following Chernozhukov et al. (2013), the asymptotic theory of this paper aims at establishing the convergence of the continuum of these binary regressions.

Note that no additional assumptions are needed to establish the consistent estimation of counterfactual distributions. Assumptions 1 to 4 are sufficient for the application of the functional delta method on $\beta_n(t)$. Similarly, other functionals of the multivariate CDF such as averages, quantile functions or variance-covariance matrices will be consistently estimated. Corollary 1 in the following section establishes the applicability of the functional delta method, section 3.3 then builds on this result and formally introduces the counterfactual framework.

3.2 Results

Based on the assumptions in the previous section, this section presents the main theoretical results. First, Theorem 1 establishes consistency, convergence, asymptotic normality and equicontinuity of the MDR estimator. Thereafter, Theorem 2 is concerned with the validity of the bootstrap. Having established the validity of the MDR estimator, Corollary 1 states that functionals of the estimator are consistently estimated. In particular, this includes counterfactual distributions in the spirit of Chernozhukov et al. (2013). The corresponding proofs can be found in Appendix A.

Theorem 1: **Asymptotic Distribution** *Let Assumptions 1 to 4 hold. Then, the MDR estimator* $\hat{\beta}_n(t)$ *of* $\beta_0(t)$ *in equation (8) satisfies* $\sqrt{n}(\hat{\beta}_n(t) - \beta_0(t)) \rightsquigarrow \dot{\Psi}^{-1}_{\beta_0,t} Z$, $\dot{\Psi}^{-1}_{\beta_0,t}$ *is the inverse of the Fréchet derivative at* $\beta_0(t)$, $Z \in \ell^\infty(\mathcal{T})$ *is the tight, mean zero Gaussian limiting distribution of* $\sqrt{n}(\Psi_n - \Psi)(\beta_0, t)$.

Theorem 1 states the main result of the theoretical analysis by establishing the asymptotic behaviour of the MDR estimator. Remarkably, the proof of Theorem 1 includes an argument showing that $\{\psi_{\beta,t} : ||\beta(t) - \beta_0(t)|| < \delta, t \in \mathcal{T}\}$ is a Donsker class. This allows for the application of a wide range of theoretical results (e.g. see Kosorok, 2008, Chapter 8.4). Further, note that the proof of Theorem 1 provides sufficient conditions for $\dot{\Psi}^{-1}_{\beta_0,t}$ to be smooth and invertible at $\beta_0(t)$ such that the asymptotic distribution is well defined. In particular, $\beta \mapsto \Psi(\beta, t)$ is shown to be Fréchet-differentiable which implies Hadamard differentiability.

Corollary 1 : Applicability of the Fuctional Delta Method *Consider the MDR estimator of $\beta_0(t)$ defined in equation (8). Recall, that $\Psi(\beta, t) : \Theta \mapsto \mathbb{L}$ where $\Theta = \mathbb{R}^{d_K}$ and the norms of $\Theta \mapsto \mathbb{L}$ are $||\cdot||$ and $||\cdot||_{\mathbb{L}}$, respectively. As a result of Theorem 1, $\sqrt{n}(\hat{\beta}_n(t) - \beta_0(t)) \rightsquigarrow \dot{\Psi}^{-1}_{\beta_0,t}Z$ with $Z \in \ell^\infty(\mathcal{T})$ being the tight process. Let $\phi : \Theta_\phi \subset \Theta \mapsto \mathbb{L}$. By Theorem 2.8 in Kosorok (2008), for any ϕ which is Hadamard-differentiable, it holds that $\sqrt{n}(\phi(\hat{\beta}_n(t)) - \phi(\beta_0(t))) \rightsquigarrow \dot{\phi}_{\beta_0(t)}(\dot{\Psi}^{-1}_{\beta_0,t}Z)$. By Theorem 2.9 in Kosorok (2008), the bootstrap applies to $\phi(\hat{\beta}_n(t))$.*

Corollary 1 establishes that functionals of the estimated CDF are consistently estimated as long as they are Hadamard-differentiable. Further, the bootstrap is applicable to these functionals too. Note that if ϕ is chosen to be Ψ, it directly follows that ϕ satisfies the requirements of Corollary 1. This implication will be useful for the consistency of counterfactual distributions. Whilst it is possible to draw inference based on the asymptotic variance derived in Theorem 1, the following result establishes the validity of the bootstrap. For the practitioner, this may be valuable tool as bootstrap techniques are flexibly implemented. Theorem 2 establishes the validity of the multiplier and the multinomial bootstrap.

Theorem 2: Validity of the Bootstrap *Let Assumptions 1 to 5 hold. Denote the non-parametric or multiplier bootstrapped estimator of $\hat{\beta}_n$ by $\hat{\beta}^\circ_n$. Then, the MDR estimator $\hat{\beta}_n(t)$ of $\beta_0(t)$ in equation (8) satisfies $\sqrt{n}(\hat{\beta}^\circ_n - \hat{\beta}_n) \rightsquigarrow k_0 Z$, where $Z \in \ell^\infty(\mathcal{T})$ is the tight, mean zero Gaussian limiting distribution of $\sqrt{n}(\Psi_n - \Psi)(\beta_0, t)$, $k_0 = 1$ for the multinomial bootstrap, $k_0 = \frac{\tau}{\mu}$ in the case of the multiplier bootstrap and where xi_i are i.i.d. weights with $\mu = E\xi_1$ and $\tau = var(\xi_1)$.*

Based on the result in theorem 2, I provide an algorithm to obtain valid confidence bands for the CDF. Algorithm 1 below is closely related to algorithm 1 in Chernozhukov et al. (2019) and outlines how confidence bands can be obtained in applied settings.

Algorithm 1 : Confidence Bands for joint CDF

1. Draw many bootstrap samples of the data indexed by $j = 1, ..., B$. Use either the multinomial or multiplier bootstrap as outlined in theorem 2.

2. For each draw, obtain an estimate of the joint CDF $\hat{F}^j_{Y_i|X_i}(t|X_i)$.
3. For each $t \in \mathcal{T}$, compute the robust standard errors by

$$\hat{s}(t) = (\hat{Q}(.75, t) - \hat{Q}(.25, t))/(\Phi^{-1}(.75) - \Phi^{-1}(.25)),$$

   where $\hat{Q}(\alpha, t)$ is the empirical α-quantile of the bootstrap sample of the CDF $\hat{F}^j_{Y_i|X_i}(t|X_i)$ at t. Φ^{-1} denotes the inverse of the standard normal distribution.

4. Define the critical value to be

$$c(p) = p\text{-quantile of } \left\{ \max_{t \in \mathcal{T}} |\hat{F}^j_{Y_i|X_i}(t|X_i) - \hat{F}_{Y_i|X_i}(t|X_i)|/\hat{s}(t) \right\},$$

   where $\hat{F}_{Y_i|X_i}(t|X_i)$ denotes the point estimate of the CDF at t.
5. Construct the confidence bands for the CDF as

$$[L'(t), U'(t)] = \left[\hat{F}_{Y_i|X_i}(t|X_i) \pm c(p)\hat{s}(t) \right].$$

To draw uniform inference on functionals of the estimated CDF, I propose to use an analogue procedure as outlined in algorithm 3 in Chernozhukov et al. (2013). Finally, section A.3 in Appendix A introduces how one can test directly whether the coefficients are constant across the univariate distributions.

3.3 Counterfactual Distributions

This section formalizes the framework outlined in section 2 and builds on the previously derived results. In general, counterfactual distributions provide a flexible and tractable tool to analyze how regressors affect the joint distributions of the outcomes. While it is feasible to draw conclusions based on the estimated coefficients in (8), they depend on the link function, which complicates the interpretation. Instead, counterfactual distributions directly connect to potential changes in the multivariate CDF. As outlined in Chernozhukov et al. (2013, section 2.2), there are three types of counterfactuals: (i) one can either modify the covariate distribution, (ii) the conditional distribution, or (iii) both. This is done by using the conditional distribution of subgroup I, $F_{Y_I|X_I}(t|X)$ which is modelled by (8), and integrating it over the covariate distribution of subgroup J, $F_{X_J}(X)$. Thus, $F_{Y\langle I|J\rangle}(t)$ represents the conditional distribution of subgroup I assuming they would have the characteristics of subgroup J. Formally, this is

$$F_{Y\langle I|J\rangle}(t) = \int_{\mathcal{X}_J} F_{Y_I|X_I}(t|X) dF_{X_J}(X). \tag{10}$$

Crucially, it has to hold that $X_J \subseteq X_I$, i.e., the support of covariates for subgroup J includes the support of covariates for subgroup I. First, consider modifications of the covariate distribution. In the simplest case, one may abstract from different subgroups and only be interested in a unit change of a specific covariate. In this context, the assumption on the common support can be dropped because the conditional CDF in (10) is integrated over all observation. The case of a binary treatment outlined in section 2 falls into this category too.

Next, consider changing the conditional distribution, i.e., the second type of counterfactuals. In practice, one takes the estimated coefficients for subgroup I and the covariate distribution of subgroup J and plug them into the model derived in equation (8). Intuitively, this answers, "How would the CDF of subgroup J look like if the regressors had the same effects as for subgroup J?" In each of these cases, the counterfactual CDF is a functional of $\hat{\beta}_n(t)$. More precisely, the estimator of $F_{Y\langle I|J\rangle}(t)$, $\hat{F}_{Y\langle I|J\rangle}(t)$, can be written as $\int_{\mathcal{X}_J} \Lambda\left[X_i'\hat{\beta}_{n,I}(t)\right] dF_{X_J}(X)$. If $\Lambda[\cdot]$ is is Hadamard-differentiable, then $\int_{\mathcal{X}_J} \Lambda[\cdot]\, dF_{X_J}(X)$ satisfies the requirements of Corollary 1. Further, Hadamard-differentiability of the link functions in assumption 2 is shown to hold in the Proof of Theorem 1.[6] Thus, counterfactual distributions are consistently estimated by the sample analog of equation (10), and the bootstrap is a valid tool to conduct inference. Similarly, note that for univariate quantile functions or conditional ones, as in equation (7), the monotone rearrangement estimator by Chernozhukov et al. (2010) is applicable.

4 Simulation Studies

The simulation studies presented in this section serve two purposes. In the first step, fixing the regression coefficients upfront, the MDR estimator is shown to converge to the true values and follow the expected Gaussian process. Then, to highlight the flexibility of MDR, I compare the estimator's performance to conditional copula models. This exercise aims to demonstrate an underlying bias-variance tradeoff: Due to their parametric assumptions, the copula models benefit from lower variances when approximating the conditional distribution. Yet, in the presence of more evolved distributions, the MDR estimator accurately accounts for non-linear effects resulting in a lower bias. For both

[6]Note that the proof of Theorem 1 includes showing that the listed link functions are Fréchet-differentiable which is an even stronger statement than Hadamard-differentiability.

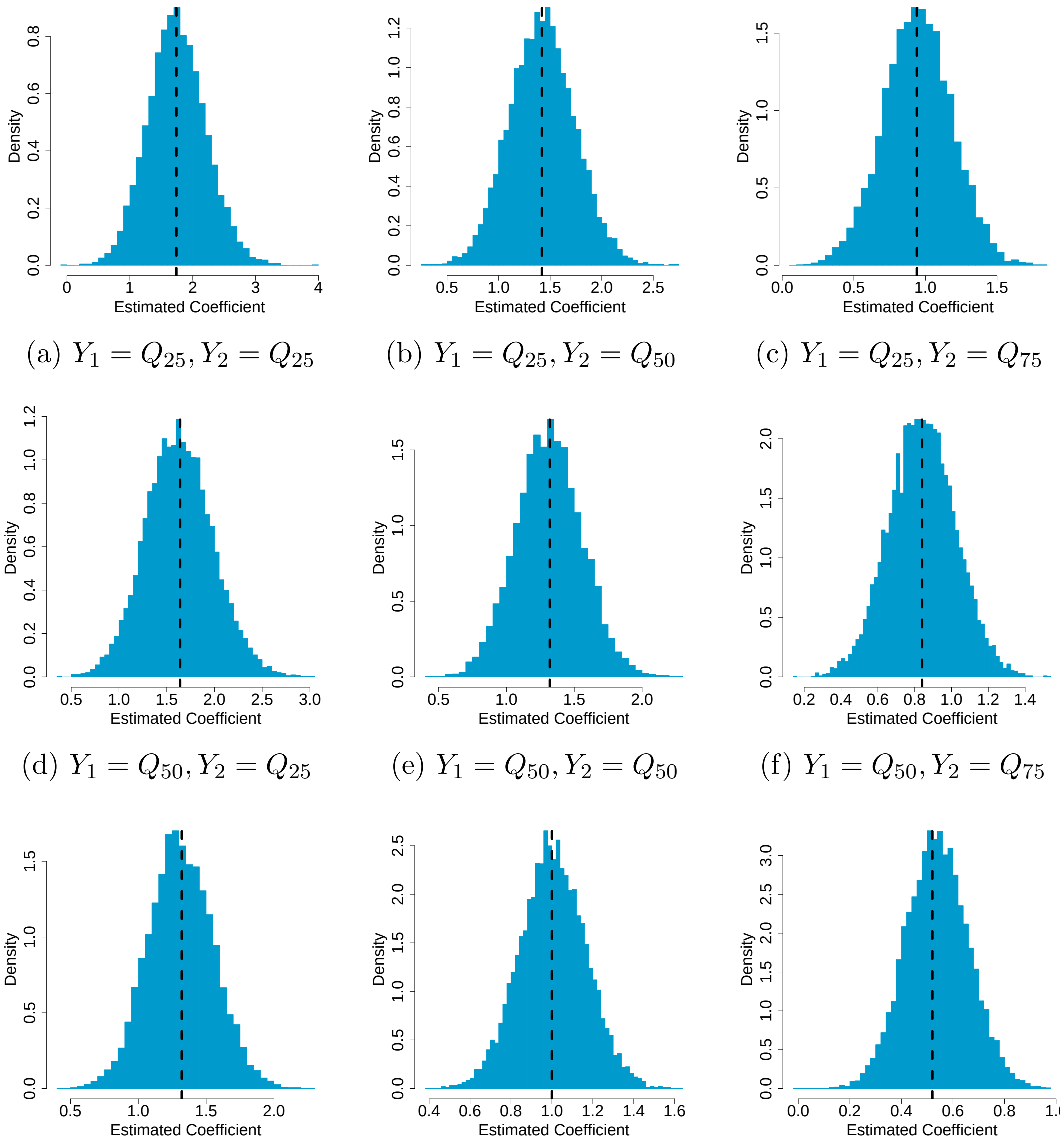
purposes, it suffices to draw data from simplistic bivariate distributions. Throughout, all distributions will depend only on one uniformly distributed regressor. Further, all simulations are based on three sample sizes $n = \{400, 1600, 6400\}$ with $10'000$ Monte Carlo experiments each. Finally, note that the results are obtained using the `CondCopulas` (Derumigny, 2020) package in `R`.

4.1 Coefficient Process

First, I consider the finite sample properties of the estimated regression coefficient process. The coefficients of the constant $\beta_{1,s1}$ have been chosen such that both outcome variables, Y_{1,s_1} and Y_{2,s_1}, take on values between 0 and 1. The uniformly distributed regressor X_{s1} affects the outcomes through a regression coefficient $\beta_{X,s1} \in (-2, 2)$. The exact specifics of the data generating process are visualized in figure 5 in Appendix B. For $n = 1600$, figure 3 presents the estimated and true coefficients of $\beta_{X,s1}$ at the 25th, 50th and 75th percentile of Y_{1,s_1} and Y_{2,s_1} respectively, resulting in nine locations of the joint distribution. From figure 3 we observe the following: (i) all estimates are centered around the true values irrespective of the location, and (ii) the distribution of estimates seems to follow a Gaussian distribution. Further, note that the variance of the estimates varies across the considered nine locations of the distribution by construction of the DGP (can be seen in figure 5), not due to a varying precision of the estimator. Table III in Appendix B contains the according mean squared errors, standard deviations and biases and supports that the MDR estimator is consistent. Finally, the same table suggests that the estimator converges at the expected rate of $\sqrt{n}$.[7]

[7]Table IV in Appendix B contains the analogue results for the estimated coefficients of the constant. All insights remain unchanged.

Figure 3: Simulation 1, Estimated Coefficients $\hat{\beta}_{X,s1}$



(a) $Y_1 = Q_{25}, Y_2 = Q_{25}$ (b) $Y_1 = Q_{25}, Y_2 = Q_{50}$ (c) $Y_1 = Q_{25}, Y_2 = Q_{75}$

(d) $Y_1 = Q_{50}, Y_2 = Q_{25}$ (e) $Y_1 = Q_{50}, Y_2 = Q_{50}$ (f) $Y_1 = Q_{50}, Y_2 = Q_{75}$

(g) $Y_1 = Q_{75}, Y_2 = Q_{25}$ (h) $Y_1 = Q_{75}, Y_2 = Q_{50}$ (i) $Y_1 = Q_{75}, Y_2 = Q_{75}$

Notes: This figure presents the estimated coefficients $\hat{\beta}_{X,s1}$ over 10'000 Monte carlo replications at 9 locations in the distribution. The sample size for each draw is 1600.

4.2 Bias-Variance Tradeoff

In the next step, I compare the performance of the MDR estimator to conditional copula models. In the following, I consider the parametric and non-parametric copula model specifications in Derumigny and Fermanian (2017) which are implemented in the R package `CondCopulas` (Derumigny, 2020).[8] In the present case, it suffices to draw data from a mean-zero normal distribution where only the standard deviation and the correlation depend on the uniformly distributed regressor. In all cases, the correlation $\rho_{Y_{1,s2},Y_{2,s2}} = \sin(4\pi \cdot X)$. Further, the standard deviations are assumed to be either both (i) 1, (ii) $(X-a)^2$ or (iii) $(X-b)^4$, where a and b are chosen such that $Y_{1,s2}$ and $Y_{2,s2}$ have, on average, a standard deviation of 1.

The estimated models require different sorts of assumptions. While for the MDR estimator only a link function has to be chosen, both copula approaches require choosing a bandwidth and either an assumption on the parametric form of the copula or the kernel. For the benchmark case presented in table I, I consider the "correctly" specified copula models assuming a normal parametric copula and a Gaussian kernel in the non-parametric case. Further, I use a bandwidth of .15 for all models. Table V in Appendix B performs various robustness checks regarding the choice of the link function, the parametric form of the copula, and the kernel. The copula models seem to depend more strongly on the respective assumptions, yet, the main insights remain unchanged. For the MDR estimator, I use a grid of 25×25 thresholds, the nonparametric copula estimator employs a grid of 25×25 pseudo observations and both copula models predict the CDF based on a grid of 25×25 values.

In contrast to the MDR estimator, the copula models do not provide a regression coefficient process. Thus, the performances are evaluated directly at the estimated CDF using an integrated mean square error (IMSE) criterion which is defined as

$$\text{IMSE} = \frac{1}{C}\sum_{j}^{C}\frac{1}{N}\sum_{i}^{N}(\hat{F}_{Y_{1,s2},Y_{2,s2}|X}(t|X)_{i,j} - F_{Y_{1,s2},Y_{2,s2}|X}(t|X)_{j})^2, \tag{11}$$

where C is the number of cells which is used to estimate the CDF, N is the number of observations in a cell j, $\hat{F}_{Y_{1,s2},Y_{2,s2}|X}(t|X)_{i,j}$ is estimated CDF of $Y_{1,s2}$ and $Y_{2,s2}$ for one observation at threshold t given X in cell j and $F_{Y_{1,s2},Y_{2,s2}|X}(t|X)_j$ is the average true value of the corresponding CDF in cell j. Assuming that the true value is constant in each cell, the $IMSE$ can be decomposed in a variance and bias term for each cell. Then, I will take the average over all cells to compare the performances. Note that implicitly, the IMSE in equation (11) weights all $25 \times 25 = 625$ cells equally, resulting in a comprehensive performance measure. As the IMSE in equation (11) corresponds to an average over many cells, the convergence rate is no longer $\sqrt(n)$. Finally, note that once we consider more data points, the standard deviation might increase as it is measured relative to the

[8] The nonparametric estimator is defined on page 158 above equation (4) whereas the parametric estimator $\hat{\theta}$ is specified on page 163. Further, note that the latter is not parametric in the strict sense as the estimation involves kernel smoothing.

cells' average.

Table I: Simulation 2, Integrated MSE, SD and Bias of $\hat{F}_{Y_{1,s2},Y_{1,s2}}(t)$

		DGP 1			DGP 2			DGP 3		
n	Model.	MSE	SD	Bias	MSE	SD	Bias	MSE	SD	Bias
400	MDR	0.0065	0.0165	0.0789	0.0080	0.0341	0.0826	0.0103	0.0394	0.0936
	NC	0.0045	0.0147	0.0657	0.0104	0.0153	0.1009	0.0166	0.0150	0.1281
	PC	0.0027	0.0211	0.0471	0.0086	0.0230	0.0897	0.0146	0.0224	0.1189
1600	MDR	0.0037	0.0252	0.0552	0.0063	0.0525	0.0598	0.0089	0.0572	0.0752
	NC	0.0027	0.0234	0.0460	0.0070	0.0226	0.0803	0.0131	0.0206	0.1126
	PC	0.0023	0.0373	0.0302	0.0065	0.0371	0.0717	0.0122	0.0331	0.1055
6400	MDR	0.0024	0.0300	0.0381	0.0055	0.0634	0.0391	0.0078	0.0665	0.0583
	NC	0.0018	0.0284	0.0320	0.0047	0.0269	0.0629	0.0106	0.0247	0.1002
	PC	0.0025	0.0454	0.0219	0.0052	0.0435	0.0579	0.0107	0.0392	0.0957

Notes: This table lists the average MSE, SD and Bias of $\hat{F}_{Y_{1,s2},Y_{2,s2}}(t)$ for simulation 2. Three models have been estimated: The Multivariate distribution regression (MDR), a non-parametric copula model (NC) and a parametric copula model (PC). The estimated CDF is evaluated at a grid of 25×25 points defined by the quantile values of $Y_{1,s2}$ and $Y_{1,s2}$. The DGPs vary according to the assumend standard deviation. All DGPs impose a mean-zero normal distribution with $\rho_{Y_{1,s2},Y_{2,s2}} = \sin(4\pi \cdot X)$. Further, DGP 1 assumes no heterogeneity ($\sigma_{Y_{1,s2}} = \sigma_{Y_{2,s2}} = 1$) where DGP 2 and 3 impose $\sigma_{Y_{1,s2}} = \sigma_{Y_{2,s2}} = (X - .378)^2$ and $\sigma_{Y_{1,s2}} = \sigma_{Y_{2,s2}} = (X - .273)^4$, respectively. The results were obtained using 10'000 Monte Carlo replications.

The results of the second simulation study are presented in table I. Considering DGP 1, the considered estimators seem to provide fairly similar results. In small samples, the MDR estimator performs slightly worse, likely because the effect of X on the distribution is overestimated. Moving on to DGP 2, the underlying bias-variance trade-off becomes obvious. Compared to the parametric copula estimator, MDR reduces the bias by 16.5 % at a cost of a 30% increase of the standard deviation ($n = 1600$). As the bias makes up for a larger share of the MSE in the case of the parametric copula estimator, both estimators provide an almost identical performance overall (MSE of 0.0063 (MDR) and 0.0065 (PC), respectively). When the complexity of the DGP is increased further, the trade-off becomes even more apparent. In particular, considering the same estimators with $n = 1600$, the MDR increases the standard deviation by 72% but reduces the bias by 29%. As the bias is roughly three times the size of the standard deviation, this results in a reduction in the MSE of 27%, a considerable improvement. Note that the chosen DGPs are still simplistic, considering stronger non-linearities would cause the MDR estimator to overperform even by a larger margin. Finally, note that the extent of the reduction varies across DGPs and sample sizes, but the underlying trade-off remains the same.

5 Application to Household Labor Supply

The division of labor within households has been studied for a long time (e.g. Shelton and John, 1996; Bianchi et al., 2000; Fuwa, 2004) and relates to intra-household decision making (e.g. Ashraf, 2009). The traditionally asymmetric distribution of labor supply may depend on childbirth (Kleven et al., 2019), tax incentives (Borella et al., 2019), norms (Evertsson, 2014), education, bargaining power (Moeeni, 2019) as well as many other factors. In the following, I consider the labor supply adjustments of spouses in response to their partner's newly receiving DIB. Many previous studies find that receiving DI payments reduces the labor earnings for the handicapped (Autor et al., 2016; Marie and Castello, 2012; Leisibach et al., 2018). In the same spirit, losing DI eligibility and the corresponding payments increase household earnings (Deshpande, 2016). In contrast, Autor et al. (2019) find no reduction in earnings for the disabled but a significant increase in spousal labor supply. Likely, the latter response intends to secure the economic stability of the household as family labor supply may act as insurance to persistent wage shocks (Blundell et al., 2016). Note that the response might depend on the complementarity of leisure time and the substitutability of childcare (Blundell et al., 2018). Lee (2020) finds substantially lower labor supply responses, mainly because spouses spend many hours caring for their disabled partners. For three reasons, MDR is a well-suited tool to analyze the bivariate distribution of spouses' earnings. First, labor earnings are known to positively correlate across households which is essential as for independent variables, univariate approaches suffice. Second, MDR can deal with the potential mass point at zero, which frequently occurs in labor income data. Finally and most importantly, receiving DIB is likely to heterogeneously affect households as low-income individuals may have fewer funds to deal with the shock. The literature suggests that spouses respond to their partners receiving DIB. In particular, I will examine whether the response depends on the income of the newly disabled partner. Intuitively, spouses of low-income individuals may need to increase the labor supply more strongly as the household has relatively fewer funds.

5.1 Data and Identification

Subsequently, I will rely on annual tax data from nine Cantons in Switzerland from 2011 to 2015. The records are provided by the Federal Social Insurance Office and are linked to social security data containing information on DI rents.[9] When filing taxes, households must provide information on the principal and their partner. In the following, I will refer

[9] https://www.bsv.admin.ch/bsv/en/home/publications-and-services/forschung/forschungsbereiche/WiSiER.html

to these as *main earner* and *second earner*, respectively. I restrict the sample only to include main earners that (i) started receiving DIB somewhere between 2011 and 2015, (ii) are married, (iii) are between 25 and 65 years old, (iv) live in a household where at least one person is working and do not divorce or die between 2011 and 2015. This leaves 30,604 observations in the estimation sample, with about half of these being treated. I define the latter to be the disability degree once a main earner newly receives DIB (no more payments imply being untreated). To account for potential confounders, I include a rich set of regressors: a linear and quadratic term for the age of the main earner, the gender of the main earner and whether the main earner holds Swiss citizenship, the number of children in the household, healthcare and drug costs, costs for already existing handicaps, the net wealth of the household as well as region and time fixed effects. Table VI and figure 6 in Appendix C present descriptive statistics of the data and a histogram of the household's labor earnings, respectively. From figure 6 it is apparent that a substantial share of main and second earners gain no income, which implies a mixed distribution.

For the results presented hereafter to have a causal interpretation, unobservables need to be uncorrelated to the degree of disability. Essentially, this would require a CIA to hold at each threshold. While disability is non-random in general, four arguments support this claim in the present case. First, the control group captures all individuals that are not yet treated but will once be. Thus, I account for unobservables that affect the likelihood of being treated. Second, the model contains many potential confounders, i.e., age, gender, health costs, and children. As receiving DIB is correlated with these socio-demographic factors, including these variables is crucial for identifying the treatment effect. In a similar setting, French and Song (2014) find that the difference between ordinary least square and instrumental variable estimates is negligible once covariates are included. Third, I exclude households where both spouses receive DIB. This ensures that a common shock does not jeopardize the treatment effect on both spouses. Finally, institutional features introduce considerable randomness in determining who will receive payments. A law passed in 2012 assigns the claims randomly to the authorized institutions, similar to the Norwegian setting (Dahl et al., 2014; Autor et al., 2019).[10] Generally, the physicians granting the DI requests differ substantially concerning their leniency (Barth et al., 2017). In fact, for Switzerland, the average approval rate of DI claims ranges from 22% to 58%, depending on the advisory institution.[11] Consequently, this law introduces further

[10]The Article 72 of the IVV (*Verordnung über die Invalidenversicherung*) instructs the officials to assign the case to one of 29 advisory institutions randomly. These institutions have to meet the requirements of the case, such as employing physicians speaking the language of the request or having the demanded physician specialty. Teams of social workers, physicians, lawyers, and administrative personnel will then coordinate decisions on DI claims.

[11]The weighted average of the approval rate is 44% (BSV, 2014). These are the latest statistics for 11 of the 29 advisory institutions. Several institutions do not publish those statistics, although they are

exogenous variation into the process. Due to these institutional factors, individuals can hardly anticipate whether and when their DI request is approved. In summary, the treatment will likely be quasi-randomly assigned once controlled for covariates.

5.2 Targeted Functionals

Following up on the leading example introduced in section 2, I propose to analyze the effect of DIB on earnings by looking at the two counterfactual distributions implied by the treatment. Thus, after having obtained the regression coefficients, I compute the CDF for the average degree of disability (74%) and no treatment (0%) based on equation (3). Further, I allow for a different treatment effect in the first year of being treated. Note that the support condition is fulfilled as we observe all values of X_i for both treated and non-treated individuals. Comparing the counterfactuals and their implied characteristics may be viewed as a ceteris paribus change of the treatment. It is instructive to look at the conditional quantile function of the second earner, provided that the main earner is in a specific range of the earnings distribution.[12] In the most trivial case, it is feasible to condition the second earners' distribution on whether the main earner has a low or high income, i.e., $Q_1(Y_1) \leq Y_1 < Q_{50}(Y_1)$ and $Q_{50}(Y_1) \leq Y_1 < Q_{100}(Y_1)$, thereby making use of equation (7). Taking the left inverse of the resulting CDFs yields the conditional QF. Finally, I define the conditional quantile treatment effect (QTE) to be the difference between the respective QFs:

$$\Delta_L(t) = F^{\leftarrow}_{Y_2\langle\cdot|Q_1(Y_{1,i})\leq Y_{1,i}<Q_{50}(Y_{1,i}),D=1\rangle}(t) - F^{\leftarrow}_{Y_2\langle\cdot|Q_1(Y_{1,i})\leq Y_{1,i}<Q_{50}(Y_{1,i}),D=0\rangle}(t) \text{ and} \tag{12}$$

$$\Delta_H(t) = F^{\leftarrow}_{Y_2\langle\cdot|Q_{50}(Y_{1,i})\leq Y_{1,i}<Q_{100}(Y_{1,i}),D=1\rangle}(t) - F^{\leftarrow}_{Y_2\langle\cdot|Q_{50}(Y_{1,i})\leq Y_{1,i}<Q_{100}(Y_{1,i}),D=0\rangle}(t). \tag{13}$$

The analog functional can be computed using the estimator by Wang et al. (2022). Further, I also employ a univariate DR approach, including a dummy variable for the high income of the main earner for comparison. I use the complementary log-log link function and a grid of 20×20 equidistant thresholds. The standard errors are obtained using 250 bootstrap draws.

5.3 Empirical Results

On average, treated main earners experience a substantial drop in earnings of roughly 50,930 CHF (32,971 CHF using OLS) from the second year of treatment onwards (table

obliged to.

[12]Note that it is feasible to look at the CDF of the second earner at exactly one quantile value of the main earners earnings as this would involve computing the multivariate pdf. However, the latter does not converge at the rate of $\sqrt{n}$ and would thus require additional theoretical results in the spirit of Rothe and Wied (2020).

VII in Appendix C). In contrast, the effect on the second earner is far smaller, with a reduction of about 934 CHF (973). Considering the main earner, this result confirms existing findings for Spain and the US (Marie and Castello, 2012; Autor et al., 2016; Gelber et al., 2017). Further, on average, second earners do not compensate for the loss in household income as suggested by Autor et al. (2019). Possibly, instead of raising the labor supply, second earners spend more time caring for their spouses (Lee, 2020). Another explanation could be that the social security system insures the average household; thus, no compensation is needed. In either case, average effects are potentially misleading as DIB may be of greater importance for low-income individuals. To see this, note that in Switzerland DIB depend on previous income only up to a degree (Leisibach et al., 2018, figure on p. 49).

Panel (a) and (b) of figure 4 present the QTE for second earners conditional on whether the main earner earnings are below or above the median (equation (12) and (13)) for MDR and the estimator by Wang et al. (2022), respectively. Panel (c) shows the results of a univariate DR approach using an additional covariate for an above-median income of the main earner. First, considering panel (a), we observe that second earners of low-income main earners increase their labor supply while the opposite holds for second earners of high-income partners. Likely, the former need to work more to secure the household's financial stability. Assuming that for these households, the shock is relatively more severe, this is in line with findings regarding other health shocks such as fatalities (Fadlon and Nielsen, 2021). Further, we observe that high-income second earners experience the highest TEs, irrespective of whether their partners are of low or high earnings. Potentially, high-income second earners are better educated and are currently working. Both would facilitate changing working hours due to a better labor market position. The heterogeneity displayed in panel (a) highlights that the response crucially depends on the location of the bivariate distribution. Turning to panels (b) and (c), it is apparent that neither of the other estimators can pick up a similar result. Concerning the estimator by Wang et al. (2022), it might be that the dependence structure is not separable in two univariate DRs. Note that while the dependence structure seems restricted by this estimator, the univariate QTE are almost identical to the ones estimated by MDR (figure 7 in Appendix C). Finally, regarding the univariate DR approach, the unconditional location of the main earner's income seems less informative than the conditional one.

Considering all the evidence, the present application confirms previous findings and contributes to the literature by analyzing the heterogeneity of the DIB's effects on earnings. In particular, receiving DIB heterogeneously affects earnings along two lines. (i) Spouses of low-income principals increase their labor supply, but the contrary is true for spouses of high-income principals. (ii) High-income spouses adapt their labor supply more strongly, likely because their better labor market position gives them more choices. In this regard, the findings of Autor et al. (2019) are confirmed and shown to depend on the level of earnings. Further, the results relate to the ongoing debate on how the poor are exposed to economic shocks and the marginal propensity to consume (e.g. Blundell et al., 2016; Misra and Surico, 2014; Kaplan et al., 2020).

Figure 4: Quantile Treatment Effects: Second to sixth year

(a) MDR (b) Wang et al. (2022) (c) Univariate DR

Notes: Panels (a)-(c) present quantile treatment effects implied by the difference of the two counterfactual distributions (average disability degree/no disability). According to equation equation (12) and (13), panel (a) shows QTE on the second earners earnings conditional on whether the main earner has an income below or above the median. Panel (b) presents the analogue result for the joint distribution implied by the estimator of Wang et al. (2022). Panel (c) shows the results of a univariate DR approach where a dummy variable for low/high income of the ME was included. The dotted lines represent uniform 95%-confidence bands computed according to Algorithm 1 in section 3 with 250 bootstrap draws. The number of observations is 30,604.

6 Conclusion

The present paper introduces a novel tool to analyze the multivariate distribution of outcomes. Compared to similar approaches in the literature, the practical advantages of MDR are threefold. First, upto date MDR is the most flexible approach to modeling a joint distribution in the presence of many covariates. While non-parametric estimators might suffer from the curse of dimensionality, other approaches restrict the distribution by either parametric assumptions like copulas or the way the dependence is modeled (Wang et al., 2022). Allowing for a flexible dependence structure might be crucial to reduce the misspecification risk and capture heterogeneous treatment effects. Second, MDR is robust as it does not require tuning parameters, nor is it sensible to the choice of the grid. Finally, MDR provides a natural framework to study how regressors affect the distribution of outcomes. Distributional treatment effects may be defined in a flexible manner. This is beneficial for numerous research questions like studying stock market indices, intergenerational mobility, or the allocation of household labor. Regarding the latter, MDR can deepen established findings by analyzing specific parts of the outcome distribution, as shown in section 5.

The framework of MDR offers various avenues for future research. For instance, the theoretical results in section 3 rely on the simplifying assumption of i.i.d. data. Note that the results could be strengthened as the corresponding theorems in Kosorok (2008) hold in more general cases. Another opportunity for future work might be developing a test on the multivariate grid to check whether the covariates' effects are indeed heterogeneous. Several additional challenges could be incorporated into MDR models. In this regard, future research may tackle the endogeneity of the regressors. Further, one could extend the sample selection setting of univariate DR (Chernozhukov et al., 2018) to the multivariate case or establish the inclusion of fixed effects similar to the model in Chernozhukov et al. (2020a). Finally, the estimated multivariate CDF offers the analysis of many more interesting functionals. For instance, one could establish the convergence rate for multivariate probability density functions, thereby extending the results in Rothe and Wied (2020).

Bibliography

Aloui, Riadh, Mohamed Safouane Ben Aïssa, and Duc Khuong Nguyen, "Global financial crisis, extreme interdependences, and contagion effects: The role of economic structure?," *Journal of Banking & Finance*, jan 2011, *35* (1), 130–141.

Ashraf, Nava, "Spousal Control and Intra-Household Decision Making: An Experimental Study in the Philippines," *American Economic Review*, aug 2009, *99* (4), 1245–1277.

Autor, David, Andreas Kostøl, Magne Mogstad, and Bradley Setzler, "Disability Benefits, Consumption Insurance, and Household Labor Supply," *American Economic Review*, jul 2019, *109* (7), 2613–2654.

_ , Mark Duggan, Kyle Greenberg, and David S. Lyle, "The Impact of Disability Benefits on Labor Supply: Evidence from the VA' s Disability Compensation Program," *American Economic Journal: Applied Economics*, jul 2016, *8* (3), 31–68.

Barth, Jürgen, Wout E L de Boer, Jason W Busse, Jan L Hoving, Sarah Kedzia, Rachel Couban, Katrin Fischer, David Y von Allmen, Jerry Spanjer, and Regina Kunz, "Inter-rater agreement in evaluation of disability: systematic review of reproducibility studies," *BMJ*, jan 2017, p. j14.

Bianchi, Suzanne M., Melissa A. Milkie, Liana C. Sayer, and John P. Robinson, "Is Anyone Doing the Housework? Trends in the Gender Division of Household Labor," *Social Forces*, sep 2000, *79* (1), 191.

Bick, Alexander and Nicola Fuchs-Schündeln, "Quantifying the Disincentive Effects of Joint Taxation on Married Women' s Labor Supply," *American Economic Review*, may 2017, *107* (5), 100–104.

Blau, Francine D and Anne E Winkler, *The Economics of Women, Men and Work*, 8th edition ed., Oxford University Press, 2018.

Blundell, Richard, Luigi Pistaferri, and Itay Saporta-Eksten, "Consumption Inequality and Family Labor Supply," *American Economic Review*, feb 2016, *106* (2), 387–435.

_ , _ , and _ , "Children, Time Allocation, and Consumption Insurance," *Journal of Political Economy*, oct 2018, *126* (S1), S73–S115.

Borella, Margherita, Mariacristina De Nardi, and Fang Yang, "Are Marriage-Related Taxes and Social Security Benefits Holding Back Female Labor Supply?," jul 2019.

Bouzebda, S. and B. Nemouchi, "Central Limit Theorems for Conditional Empirical and Conditional U-Processes of Stationary Mixing Sequences," *Mathematical Methods of Statistics*, jul 2019, *28* (3), 169–207.

BSV, "SuisseMED@P-Reporting," 2014.

Callaway, Brantly, Tong Li, and Irina Murtazashvili, "Nonlinear Approaches to Intergenerational Income Mobility allowing for Measurement Error," *arXiv preprint arXiv:2107.09235*, 2021.

Canadian Press, "Trudeau and Scheer blame one another for blocking emergency disability benefit," *The Canadian Press*, Jun 2020.

Cesarini, David, Erik Lindqvist, Matthew J. Notowidigdo, and Robert Östling, "The Effect of Wealth on Individual and Household Labor Supply: Evidence from Swedish Lotteries," *American Economic Review*, dec 2017, *107* (12), 3917–3946.

Ćevid, Domagoj, Loris Michel, Jeffrey Näf, Nicolai Meinshausen, and Peter Bühlmann, "Distributional random forests: Heterogeneity adjustment and multivariate distributional regression," *arXiv preprint arXiv:2005.14458*, 2020.

CH Media, "Abschaffung der Heiratsstrafe: Ständerat lehnt Sololauf ab," Aargauer Zeitung, September 2020.

Chavas, Jean-Paul, "The dynamics and volatility of prices in multiple markets: a quantile approach," *Empirical Economics*, jan 2020.

Chernozhukov, Victor, Alfred Galichon, Marc Hallin, and Marc Henry, "Monge-Kantorovich depth, quantiles, ranks and signs," *Ann. Statist.*, 02 2017, *45* (1), 223–256.

_ and Iván Fernández-Val, "Subsampling Inference on Quantile Regression Processes," *Sankhyā: The Indian Journal of Statistics (2003-2007)*, 2005, *67* (2), 253–276.

_ , Iván Fernández-Val, and Alfred Galichon, "Quantile and probability curves without crossing," *Econometrica*, 2010, *78* (3), 1093–1125.

_ , Iván Fernández-Val, and Blaise Melly, "Inference on Counterfactual Distributions," *Econometrica*, 2013, *81* (6), 2205–2268.

_ , Ivan Fernandez-Val, and Martin Weidner, "Network and panel quantile effects via distribution regression," *Journal of Econometrics*, 2020.

_ , Iván Fernández-Val, and Siyi Luo, "Distribution regression with sample selection, with an application to wage decompositions in the UK," *arXiv preprint arXiv:1811.11603*, 2018.

_ , Iván Fernández-Val, Blaise Melly, and Kaspar Wüthrich, "Generic Inference on Quantile and Quantile Effect Functions for Discrete Outcomes," *Journal of the American Statistical Association*, jun 2019, *115* (529), 123–137.

_ , Iván Fernández-Val, Whitney Newey, Sami Stouli, and Francis Vella, "Semiparametric Estimation of Structural Functions in Nonseparable Triangular Models," *Quantitative Economics*, May 2020, *11* (2).

Chetty, Raj and Nathaniel Hendren, "The impacts of neighborhoods on intergenerational mobility I: Childhood exposure effects," *The Quarterly Journal of Economics*, 2018, *133* (3), 1107–1162.

— **and** — , "The impacts of neighborhoods on intergenerational mobility II: County-level estimates," *The Quarterly Journal of Economics*, 2018, *133* (3), 1163–1228.

— **, John N Friedman, Emmanuel Saez, Nicholas Turner, and Danny Yagan**, "Income segregation and intergenerational mobility across colleges in the United States," *The Quarterly Journal of Economics*, 2020, *135* (3), 1567–1633.

Christoffersen, Peter, Vihang Errunza, Kris Jacobs, and Hugues Langlois, "Is the Potential for International Diversification Disappearing? A Dynamic Copula Approach," *Review of Financial Studies*, oct 2012, *25* (12), 3711–3751.

Cox, D. R., "Regression Models and Life-Tables," *Journal of the Royal Statistical Society: Series B (Methodological)*, jan 1972, *34* (2), 187–202.

Dahl, Gordon B., Andreas Ravndal Kostøl, and Magne Mogstad, "Family Welfare Cultures," *The Quarterly Journal of Economics*, aug 2014, *129* (4), 1711–1752.

Delattre, Sylvain and Etienne Roquain, "On empirical distribution function of high-dimensional Gaussian vector components with an application to multiple testing," *Bernoulli*, feb 2016, *22* (1), 302–324.

der Vaart, Aad Van, *Asymptotic statistics*, Vol. 3, Cambridge university press, 2000.

Derumigny, Alexis, *CondCopulas: Estimation of conditional copula models* 2020. R package.

— **and Jean-David Fermanian**, "About tests of the "simplifying" assumption for conditional copulas," *Dependence Modeling*, aug 2017, *5* (1), 154–197.

Deshpande, Manasi, "The Effect of Disability Payments on Household Earnings and Income: Evidence from the SSI Children's Program," *Review of Economics and Statistics*, 2016, *98* (4), 638–654.

Dudley, R. M., "Weak convergence of probabilities on nonseparable metric spaces and empirical measures on Euclidean spaces," *Illinois Journal of Mathematics*, mar 1966, *10* (1), 109–126.

Eika, Lasse, Magne Mogstad, and Basit Zafar, "Educational Assortative Mating and Household Income Inequality," *Journal of Political Economy*, dec 2019, *127* (6), 2795–2835.

Evertsson, Marie, "Gender Ideology and the Sharing of Housework and Child Care in Sweden," *Journal of Family Issues*, feb 2014, *35* (7), 927–949.

Fadlon, Itzik and Torben Heien Nielsen, "Family labor supply responses to severe health shocks: Evidence from Danish administrative records," *American Economic Journal: Applied Economics*, 2021.

Fermanian, Jean-David, "Goodness-of-fit tests for copulas," *Journal of Multivariate Analysis*, July 2005, *95* (1), 119–152.

_ **and Olivier Lopez**, "Single-index copulas," *Journal of Multivariate Analysis*, may 2018, *165*, 27–55.

Foresi, Silverio and Franco Peracchi, "The Conditional Distribution of Excess Returns: An Empirical Analysis," *Journal of the American Statistical Association*, jun 1995, *90* (430), 451–466.

French, Eric and Jae Song, "The Effect of Disability Insurance Receipt on Labor Supply," *American Economic Journal: Economic Policy*, may 2014, *6* (2), 291–337.

Fuwa, Makiko, "Macro-level Gender Inequality and the Division of Household Labor in 22 Countries," *American Sociological Review*, dec 2004, *69* (6), 751–767.

Gelber, Alexander, Timothy J. Moore, and Alexander Strand, "The Effect of Disability Insurance Payments on Beneficiaries' Earnings," *American Economic Journal: Economic Policy*, aug 2017, *9* (3), 229–261.

Gijbels, Irène, Noël Veraverbeke, and Marel Omelka, "Conditional copulas, association measures and their applications," *Computational Statistics & Data Analysis*, may 2011, *55* (5), 1919–1932.

Goux, Dominique, Eric Maurin, and Barbara Petrongolo, "Worktime regulations and spousal labor supply," *American Economic Review*, 2014, *104* (1), 252–76.

Halla, Martin, Julia Schmieder, and Andrea Weber, "Job Displacement, Family Dynamics, and Spousal Labor Supply," *American Economic Journal: Applied Economics*, oct 2020, *12* (4), 253–287.

Ho, Anson T. Y., Kim P. Huynh, and David T. Jacho-Chávez, "Flexible Estimation of Copulas: An Application to the US Housing Crisis," *Journal of Applied Econometrics*, jan 2015, *31* (3), 603–610.

Hyslop, Dean R, "Rising U.S. Earnings Inequality and Family Labor Supply: The Covariance Structure of Intrafamily Earnings," *American Economic Review*, sep 2001, *91* (4), 755–777.

Jung, Sin-Ho, "Regression Analysis for Long-Term Survival Rate," *Biometrika*, March 1996, *83* (1), 227–232.

Kaplan, Greg, Benjamin Moll, and Giovanni Violante, "The Great Lockdown and the Big Stimulus: Tracing the Pandemic Possibility Frontier for the U.S.," University of Chicago, Becker Friedman Institute for Economics Writing Working Paper 119 2020.

Klein, Nadja, Torsten Hothorn, Luisa Barbanti, and Thomas Kneib, "Multivariate conditional transformation models," *Scandinavian Journal of Statistics*, 2019.

Kleven, Henrik, Camille Landais, and Jakob Egholt Søgaard, "Children and Gender Inequality: Evidence from Denmark," *American Economic Journal: Applied Economics*, oct 2019, *11* (4), 181–209.

Koenker, Roger and Gilbert Bassett, "Regression Quantiles," *Econometrica*, jan 1978, *46* (1), 33.

Kosorok, Michael R., *Introduction to Empirical Processes and Semiparametric Inference*, Springer New York, 2008.

Kühhirt, Michael, "Childbirth and the Long-Term Division of Labour within Couples: How do Substitution, Bargaining Power, and Norms affect Parents' Time Allocation in West Germany?," *European Sociological Review*, apr 2011, *28* (5), 565–582.

Lavy, Victor, Assaf Kott, and Genia Rachkovski, "Does Remedial Education in Late Childhood Pay Off After All? Long-Run Consequences for University Schooling, Labor Market Outcomes, and Intergenerational Mobility," *Journal of Labor Economics*, 2022, *40* (1), 239–282.

Lee, Siha, "Spousal Labor Supply, Caregiving, and the Value of Disability Insurance," CLEF Working Paper Series 21, Canadian Labour Economics Forum (CLEF), University of Waterloo 2020.

Leisibach, Patrick, Christoph Schaltegger, and Lukas Schmid, "Arbeitsanreize in der sozialen Sicherheit," Luzern, SECO 2018.

Manoli, Day and Andrea Weber, "Nonparametric evidence on the effects of financial incentives on retirement decisions," *American Economic Journal: Economic Policy*, 2016, *8* (4), 160–82.

Marie, Olivier and Judit Vall Castello, "Measuring the (income) effect of disability insurance generosity on labour market participation," *Journal of Public Economics*, feb 2012, *96* (1-2), 198–210.

Misra, Kanishka and Paolo Surico, "Consumption, Income Changes, and Heterogeneity: Evidence from Two Fiscal Stimulus Programs," *American Economic Journal: Macroeconomics*, 2014, *6* (4), 84–106.

Moeeni, Safoura, "Married women's labor force participation and intra-household bargaining power," *Empirical Economics*, nov 2019.

Nekoei, Arash and Andrea Weber, "Does extending unemployment benefits improve job quality?," *American Economic Review*, 2017, *107* (2), 527–61.

Patton, Andrew J., "A review of copula models for economic time series," *Journal of Multivariate Analysis*, sep 2012, *110*, 4–18.

Peters, Rudi, "Steuerliche Ungleichbehandlung von verheirateten und unverheirateten Paaren in den Kantonen und beim Bund. ESTV.," 2014.

Poon, Ser-Huang, Michael Rockinger, and Jonathan Tawn, "Extreme value dependence in financial markets: Diagnostics, models, and financial implications," *The Review of Financial Studies*, 2004, *17* (2), 581–610.

Portier, François and Johan Segers, "On the weak convergence of the empirical conditional copula under a simplifying assumption," *Journal of Multivariate Analysis*, jul 2018, *166*, 160–181.

Rothe, Christoph and Dominik Wied, "Estimating derivatives of function-valued parameters in a class of moment condition models," *Journal of Econometrics*, jul 2020, *217* (1), 1–19.

Schöchli, Hansueli, "Einzelbesteuerung von Ehepartnern bahnt sich im Parlament an," Neue Zürcher Zeitung, 2019.

Schwartz, Christine R., "Earnings Inequality and the Changing Association between Spouses' Earnings," *American Journal of Sociology*, mar 2010, *115* (5), 1524–1557.

Shelton, Beth Anne and Daphne John, "The Division of Household Labor," *Annual Review of Sociology*, aug 1996, *22* (1), 299–322.

van der Vaart, Aad W. and Jon A. Wellner, *Weak Convergence and Empirical Processes*, Springer New York, 1996.

Wald, Abraham, "Note on the Consistency of the Maximum Likelihood Estimate," *The Annals of Mathematical Statistics*, 1949, *20* (4), 595–601.

Walker, Peter and Jessica Elgot, "Disability benefit change shows Tories are still 'nasty party', says Corbyn," The Guardian, 2017.

Wang, Yunyun, Tatsushi Oka, and Dan Zhu, "Bivariate Distribution Regression with Application to Insurance Data," *arXiv preprint arXiv:2203.12228*, 2022.

Ward, Zachary, "Internal Migration, Education, and Intergenerational Mobility: Evidence from American History," *Journal of Human Resources*, 2020, pp. 0619–10265R2.

White, Halbert, "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, jan 1982, *50* (1), 1.

Williams, O. Dale and James E. Grizzle, "Analysis of Contingency Tables Having Ordered Response Categories," *Journal of the American Statistical Association*, mar 1972, *67* (337), 55–63.

Zimmer, David M., "The Role of Copulas in the Housing Crisis," *Review of Economics and Statistics*, may 2012, *94* (2), 607–620.

Zuo, Yijun and Robert Serfling, "General Notions of Statistical Depth Function," *The Annals of Statistics*, 2000, *28* (2), 461–482.

Appendix

A Theoretical Results

A.1 Proof of Theorems 1 and 2

The proof of theorems 1 and 2 relies on the master theorem for Z-estimators outlined in Kosorok (2008, p. 247, theorem 13.4). Note that both theorems are direct consequences of the following six conditions of the master theorem. Thus, it suffices to verify that the conditions hold. In the following, denote the parameter space by $\Theta = \mathbb{R}^{d_K}$. In general, $\Psi(\beta, t)$ maps to a space $\mathbb{L}$ with norm $||\cdot||_{\mathbb{L}}$. In the case of MDR, I will consider the infinity norm for space $\mathbb{L}$. Further, let $\hat{\beta}_n$ be an approximate zero of Ψ_n and $\hat{\beta}_n^\circ$ be a minimizer of $\sup_{t\in\mathcal{T}} |\Psi_n^\circ(\beta, t)|$ where $\Psi_n^\circ(\beta, t) = \mathbb{P}_n^\circ \psi_{\beta,t}$, and $\mathbb{P}_n^\circ f = n^{-1}\sum_{i=1}^n \frac{\xi_i}{\bar{\xi}} f(X_i)$ denotes the non-parametric or multiplier bootstrap where $\bar{\xi} = n^{-1}\sum_{i=1}^n \xi_i$. In the following, $(...|\mathcal{X}_n)$ states that we condition on the data. By Assumption 1, $\Psi(\beta, t) = P\psi_{\beta,t}$, where P is the probability measure and $\psi_{\beta,t}$ is the derivative of the log likelihood, that is

$$\psi_{\beta,t} = (\Lambda\left[X_i'\beta(t)\right] - y_i(t)) \left(\frac{\lambda\left[X_i'\beta(t)\right] X_i}{\Lambda\left[X_i'\beta(t)\right](1 - \Lambda\left[X_i'\beta(t)\right])}\right), \tag{14}$$

where $\lambda(\cdot)$ denotes the derivative of $\Lambda(\cdot)$ and $y_i(t) = \mathbf{1}(Y_{1,i} \le t_1, ..., Y_{d,i} \le t_d)$. Thus $\Psi(\beta, t)$ can be derived by integrating over the probability measure, i.e.

$$\Psi(\beta, t) = E\left[(\Lambda\left[X_i'\beta_t\right] - y_i(t)) \left(\frac{\lambda\left[X_i'\beta_t\right] X_i}{\Lambda\left[X_i'\beta_t\right](1 - \Lambda\left[X_i'\beta_t\right])}\right)\right]. \tag{15}$$

Condition 1 : Identifiability
$\beta \mapsto \Psi(\beta, t)$ *satisfies* $||\Psi(\beta_n, t)||_{\mathbb{L}} \to 0$ *implies* $||\beta_n(t) - \beta_0(t)|| \to 0$ *for any* $\{\beta_n(t)\} \in \Theta$.

Condition 2 : Glivenko-Cantelli
$\{\psi_{\beta,t}; \beta \in \Theta, t \in \mathcal{T}\}$ *is P-Glivenko-Cantelli.*

Condition 3 : Donsker Class
F_δ *in* $F_\delta \equiv \{\psi_{\beta,t} : ||\beta(t) - \beta_0(t)|| < \delta, t \in \mathcal{T}\}$ *is P-Donsker for some* $\delta > 0$.

Condition 4 : Equicontinuity
$\sup_{t\in\mathcal{T}} P(\psi_{\beta,t} - \psi_{\beta_0,t})^2 \to 0$, *as* $\beta(t) \to \beta_0(t)$.

Condition 5 : Approximate Zeros
$||\Psi_n(\hat{\beta}_n, t)||_{\mathbb{L}} = o_P(n^{-1/2})$ *and* $P\left(\sqrt{n}||\Psi_n^\circ(\hat{\beta}_n^\circ, t)||_{\mathbb{L}} > \eta|\mathcal{X}_n\right) = o_P(1)$ *for every* $\eta > 0$.

Condition 6 : Smoothness and Invertibility of the Derivative
$\beta \mapsto \Psi(\beta, t)$ *is Fréchet-differentiable at* $\beta_0(t)$ *with continuously invertible derivative* $\dot{\Psi}_{\beta_0,t}$.

Proof. Condition 1. Using the infinity norm, identifiability is given by the fact that $||\Psi(\beta,t)||_\infty \to 0$ implies that $\Lambda\left[X_i'\beta_t\right] \to F_{Y|X}(t)$ which is $||\beta_n - \beta_0||_\infty \to 0$. Condition 2 and 3. I verify that $\{\psi_{\beta,t} : ||\beta(t) - \beta_0(t)|| < \delta, t \in \mathcal{T}\}$ is P-Donsker in with the following argument. First, note that $\{P(X_q) : q = 1, ..., d_x\}$, $\mathcal{F}_1 = \{X'\beta : \beta \in \mathbb{R}^{d_x}\}$ and $\mathcal{F}_2 = \{\mathbf{1}(Y_{1,i} \leq t_1, ..., Y_{d,i} \leq t_d) : t \in \mathcal{T}\}$ are VC classes of functions. In particular, the multivariate indicator functions, $\mathcal{F}_2$, are shown to be VC-classes by van der Vaart and Wellner (1996, example 2.6.1 and 2.10.4). Following the argument in (see Chernozhukov et al., 2013, p. 2263), $\mathcal{G} = \{(\Lambda(\mathcal{F}_1) - \mathcal{F}_2)\frac{\lambda[\mathcal{F}_1]}{\Lambda[\mathcal{F}_1](1-\Lambda[\mathcal{F}_1])}P(X_q) : q = 1, ..., d_x\}$ is a Lipschitz transformation of VC classes. The Lipschitz coefficient is bounded by const$||X||$ and envelope function const$||X||$. Further, the envelope function is square integrable. By Example 19.9 in Van der Vaart (2000), $\mathcal{G}$ is Donsker. As any Donsker class is also Glivenko-Cantelli (Kosorok, 2008, p. 19), condition 2 is fulfilled too. Condition 4. As $\beta \to \beta_0$, $\Lambda(\beta) \to \Lambda(\beta_0)$ and $\lambda(\beta) \to \lambda(\beta_0)$ for all $t \in \mathcal{T}$. Thus, $\sup_{t\in\mathcal{T}} P(\psi_{\beta,t} - \psi_{\beta_0,t})^2 \to 0$. Condition 5. Again, using the infinity norm, the sample analogue of equation (15) can be shown to converge to 0 almost surely. To see this, note that $\left|\left|\frac{1}{n}\sum_{i=1}^n \psi_{\hat{\beta}_n,t}\right|\right|_\infty \xrightarrow{a.s.} 0$ because $\Lambda\left[X_i'\hat{\beta}_n(t)\right] \xrightarrow{a.s.} y_i(t)$ as $n \to \infty$. As $\psi_{\beta,t}$ is Donsker and thus Glivenko-Cantelli, theorem 10.13 part (viii) in Kosorok (2008, p. 187) yields that $P\left(\left|\left|\frac{1}{n}\sum_{i=1}^n \frac{\xi_i}{\bar{\xi}}\psi_{\hat{\beta}_n^\circ,t}\right|\right|_\infty > \eta|\mathcal{X}_n\right) \xrightarrow{a.s.} 0$ which is equivalent to the desired statement in the second part of the condition.[13] Condition 6. This condition is verified by the fact that $\Psi(\beta,t) : \mathcal{M} \times \Theta \mapsto \Theta$ meets the conditions in Lemma E.1 and E.2 in Chernozhukov et al. (2013, p. 2254) which in turn requires assumption 3. Table II lists the parametric forms of the link functions introduced in Assumption 2. Note that $\frac{\partial}{\partial(b',t)}\Psi(b,t) = [J(b,t), R(b,t)]$. By the dominated convergence theorem, the a.s. continuity of $\frac{\partial}{\partial b'}\psi_{b,t}(Y_i, X_i)$ and $f_{Y|X}(t|X)$ and $\frac{\lambda\left[X_i'\beta_t\right]X_i}{\Lambda\left[X_i'\beta_t\right](1-\Lambda\left[X_i'\beta_t\right])}$ being uniformly bounded on $b \in R^{dx}$, all these link functions satisfy this condition (see Chernozhukov et al., 2013, p. 2263). ■

A.2 Explicit Forms of the Link Functions

The following table entails the parametric forms of $J(b,t)$ and $R(b,t)$ for the link functions introduced in section 3. These are defined as $\frac{\partial}{\partial(b',t)}\Psi(b,t) = [J(b,t), R(b,t)]$ where $\Psi(\beta,t) = E\left[(\Lambda\left[X_i'\beta_t\right] - y_i(t))\left(\frac{\lambda\left[X_i'\beta_t\right]X_i}{\Lambda\left[X_i'\beta_t\right](1-\Lambda\left[X_i'\beta_t\right])}\right)\right]$.

[13] Note that the definition of the multiplier bootstrap is slightly different than the bootstrap theorem 10.13, but it does not change the conclusions (Kosorok, 2008, p. 244).

Table II: Parametric Link Functions

Link	$\Lambda(X'\beta)$	$J(b,t)$	$R(b,t)$
Linear	$X_i'\beta$	$E\left[\left(y_i(t)\frac{1-2X_i'\beta}{(X_i'\beta_t(1-X_i'\beta_t))^2}-\frac{1}{(1-X_i\beta)^2}\right)X_iX_i'\right]$	$-E[f_{Y\|X}(t\|X)\frac{1}{X_i'\beta_t(1-X_i'\beta_t)}X_i]$
Logit	$\frac{\exp(X_i'\beta_t)}{1+\exp(X_i'\beta_t)}$	$E\left[\frac{\exp(X_i'\beta_t)}{(1+\exp(X_i'\beta_t))^2}X_iX_i'\right]$	$-E[f_{Y\|X}(t\|X)X_i]$
Probit	$\int_{-\infty}^{X_i'\beta_t}\frac{1}{\sqrt{2\pi}}\exp(\frac{1}{2}z^2)dz$	$E\Bigg[\frac{\exp(-(X_i'\beta)^2/2)}{\sqrt{2\pi}}\left(\frac{\frac{\exp(-(X_i'\beta)^2/2)}{\sqrt{2\pi}}+X_i'\beta\left(1-\int_{-\infty}^{X_i'\beta_t}\frac{1}{\sqrt{2\pi}}\exp(\frac{1}{2}z^2)dz\right)}{\left(1-\int_{-\infty}^{X_i'\beta_t}\frac{1}{\sqrt{2\pi}}\exp(\frac{1}{2}z^2)dz\right)^2}\right.$ $-y_i(t)\left(\frac{X_i'\beta\left(\int_{-\infty}^{X_i'\beta_t}\frac{1}{\sqrt{2\pi}}\exp(\frac{1}{2}z^2)dz(1-\int_{-\infty}^{X_i'\beta_t}\frac{1}{\sqrt{2\pi}}\exp(\frac{1}{2}z^2)dz)\right)-\frac{\exp(-(X_i'\beta)^2/2)}{\sqrt{2\pi}}}{(\int_{-\infty}^{X_i'\beta_t}\frac{1}{\sqrt{2\pi}}\exp(\frac{1}{2}z^2)dz(1-\int_{-\infty}^{X_i'\beta_t}\frac{1}{\sqrt{2\pi}}\exp(\frac{1}{2}z^2)dz))^2}\right.$ $\left.\left.+\frac{2\frac{\exp(-(X_i'\beta)^2/2)}{\sqrt{2\pi}}\left(\int_{-\infty}^{X_i'\beta_t}\frac{1}{\sqrt{2\pi}}\exp(\frac{1}{2}z^2)dz\right)}{(\int_{-\infty}^{X_i'\beta_t}\frac{1}{\sqrt{2\pi}}\exp(\frac{1}{2}z^2)dz(1-\int_{-\infty}^{X_i'\beta_t}\frac{1}{\sqrt{2\pi}}\exp(\frac{1}{2}z^2)dz))^2}\right)\right)X_iX_i'\Bigg]$	$-E\left[f_{Y\|X}(t\|X)\frac{\exp(-(X_i'\beta)^2/2)/\sqrt{2\pi}}{\int_{-\infty}^{X_i'\beta_t}\frac{1}{\sqrt{2\pi}}\exp(\frac{1}{2}z^2)dz\left(1-\int_{-\infty}^{X_i'\beta_t}\frac{1}{\sqrt{2\pi}}\exp(\frac{1}{2}z^2)dz\right)}X_i\right]$
Cloglog	$1-\exp(-\exp(X_i'\beta))$	$E\left[\exp(X_i'\beta)\left(1-y_i(t)\frac{1-\exp(-\exp(X_i'\beta))-\exp(X_i'\beta-\exp(X_i'\beta))}{(1-\exp(-\exp(X_i'\beta)))^2}\right)X_iX_i'\right]$	$-E\left[f_{Y\|X}(t\|X)\frac{\exp(X_i'\beta)}{1-\exp(-\exp(X_i'\beta))}X_i\right]$
Cauchit	$.5+\frac{1}{\pi}\arctan(X_i'\beta)$	$E\Bigg[\Bigg(\frac{1}{\pi}\frac{1}{1+(X_i'\beta)^2}\frac{\frac{1}{\pi}\frac{1}{1+(X_i'\beta)^2}}{.25-\frac{1}{\pi^2}(\arctan(X_i'\beta))^2}$ $+\left(.5+\frac{1}{\pi}\arctan(X_i'\beta)-y_i(t)\right)\frac{2}{\pi(1+((X_i'\beta)^2)^2}$ $\frac{\frac{1}{\pi^2}\arctan(X_i'\beta)-X_i'\beta\left(.25-\frac{1}{\pi^2}(\arctan(X_i'\beta))^2\right)}{\left(.25-\frac{1}{\pi^2}(\arctan(X_i'\beta))^2\right)^2}\Bigg)X_iX_i'\Bigg]$	$-E[f_{Y\|X}(t\|X)\frac{\frac{1}{\pi}\frac{1}{1+(X_i'\beta)^2}}{.25-\frac{1}{\pi^2}(\arctan(X_i'\beta))^2}X_i]$

Notes: $\Psi(\beta,t)=E\left[(\Lambda\left[X_i'\beta_t\right]-y_i(t))\left(\frac{\lambda\left[X_i'\beta_t\right]X_i}{\Lambda\left[X_i'\beta_t\right](1-\Lambda\left[X_i'\beta_t\right])}\right)\right]$ *and* $\frac{\partial}{\partial(b',t)}\Psi(b,t)=[J(b,t),R(b,t)]$

A.3 Testing

This section outlines a testing framework to serve three purposes. First, one may be interested in whether the effect of a specific covariate is constant across all thresholds $t \in \mathbb{R}^d$. More precisely, this is testable by setting the null hypothesis $H_0 : \hat{\beta}_j(t) = \hat{\beta}_j(Q_{50}(Y))$ and the alternative $H_1 : \hat{\beta}_j(t) \neq \hat{\beta}_j(Q_{50}(Y))$, where j identifies the regressor X_j and $\hat{\beta}_j(Q_{50}(Y))$ denotes the coefficient on X_j at the median of all outcomes Y, $Q_{50}(Y)$. Of course, other reference values than the median can be chosen. Second and with respect to counterfactual distributions, it is natural to test whether multiple CDFs are sufficiently different. While it is possible to test multivariate CDFs, I propose to directly test the marginal CDF's of the outcome. As argued by Fermanian (2005), all approaches to test the former entail certain drawbacks, in particular, they frequently require a distributional assumption. Instead, it is valid to test the marginal CDF's as these are consistently estimated. This type of tests may be executed by the well known two-sample Kolmogorov-Smirnov test. Third and in the same spirit, one may directly target summary statistics of multivariate distributions such as averages, variances or correlations. Having established that these statistics are consistently estimated, they can be tested using t-test over bootstrap draws again in the spirit of a Kolmogorov-Smirnov test.

As outlined in Chernozhukov and Fernández-Val (2005), the Kolmogorov-Smirnov tests rely on bootstrap draws to form the corresponding test statistics. To illustrate the procedure, I outline the first type of the aforementioned tests in the following. Denote the test statistic for the point estimates at threshold t by $T(t)$ and the corresponding statistic of each bootstrap draw by $T_b(t)$. To test whether the coefficient $\hat{\beta}_{n,j}$ of variable X_j is constant across the distribution, I make use of the test statistics $T(t)$ in equation (16). Note that the bootstrapped statistic in equation (17) is recentered at the average values of the point estimates.

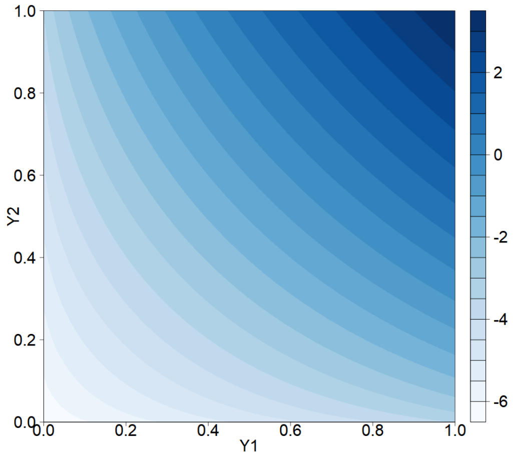
$$T(t) = \frac{\sqrt{N}\left|\hat{\beta}_j(t) - \hat{\beta}_j(Q_{50}(Y))\right|}{s.e.(\hat{\beta}_j(t) - \hat{\beta}_j(Q_{50}(Y))} \tag{16}$$

$$T_b(t) = \frac{\sqrt{M}\left|\hat{\beta}_{b,j}(t) - \hat{\beta}_{b,j}(Q_{50}(Y)) - \left(\hat{\beta}_j(t) - \hat{\beta}_j(Q_{50}(Y))\right)\right|}{s.e.(\hat{\beta}_{b,j}(t) - \hat{\beta}_{b,j}(Q_{50}(Y))}, \tag{17}$$

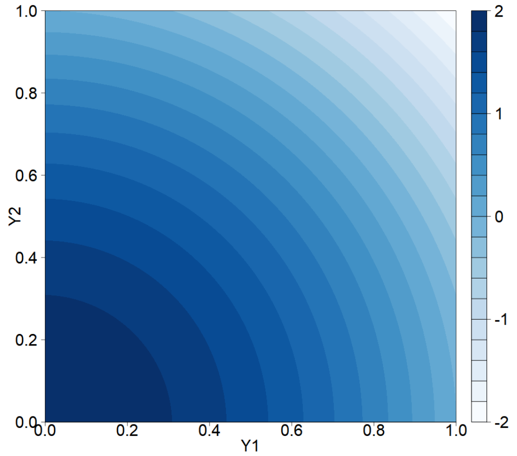
where N denotes the number of observations M is the number of observations for each draw from the bootstrap. Note that by setting $M < N$ one may reduce computation time. Further, define $T^\star = \max_{t \in \mathbb{R}^d} T(t)$ and $T_b^\star = \max_{t \in \mathbb{R}^d} T_b(t)$. Finally, the p-value is computed as the number of cases in which the bootstrapped statistic is larger than the one of the point estimates: $\frac{1}{B}\sum_{b=1}^{B} \mathrm{I}(T^\star \leq T_b^\star)$, where B is the number of bootstrap replications.

B Simulation Results

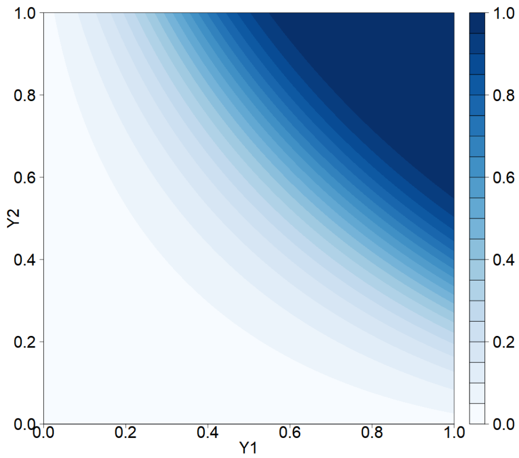
Figure 5: Simulation 1, Data Generating Process



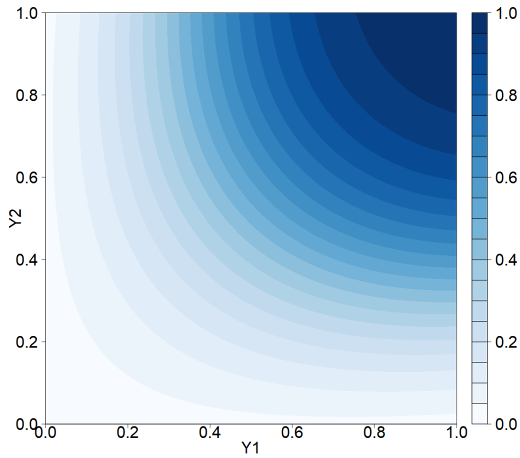
(a) $\beta_{1,s1}$



(b) $\beta_{X,s1}$



(c) $F_{Y_{1,s1},Y_{2,s1}|X=0}$



(d) $F_{Y_{1,s1},Y_{2,s1}|X=1}$

Notes: Panel (a) and (b) of this figure describe the regression coefficients for both, the constant and the uniformly distributed regressor, used to generate data for simulation 1. Panel (c) and (d) present the bivariate distribution of $Y_{1,s1}$ and $Y_{2,s1}$ provided that the regressor X is either zero or one.

Table III: Simulation 1, MSE, SD and Bias of $\hat{\beta}_{X,s1}$

		Y_1	Q_{25}			Q_{50}			Q_{75}		
n	Crit.	Y_2	Q_{25}	Q_{50}	Q_{75}	Q_{25}	Q_{50}	Q_{75}	Q_{25}	Q_{50}	Q_{75}
400	MSE		1.0017	0.4265	0.2266	0.5603	0.2422	0.1290	0.2420	0.1056	0.0621
	SD		0.9987	0.6525	0.4759	0.7480	0.4919	0.3591	0.4918	0.3247	0.2479
	Bias		0.0654	0.0285	0.0121	0.0290	0.0125	0.0084	0.0095	0.0120	0.0242
1600	MSE		0.4624	0.3175	0.2389	0.3546	0.2435	0.1805	0.2397	0.1622	0.1232
	SD		0.4624	0.3175	0.2388	0.3545	0.2434	0.1805	0.2396	0.1621	0.1218
	Bias		0.0088	0.0023	0.0060	0.0094	0.0069	0.0013	0.0064	0.0054	0.0188
6400	MSE		0.2265	0.1557	0.1180	0.1766	0.1195	0.0893	0.1208	0.0800	0.0638
	SD		0.2264	0.1557	0.1180	0.1761	0.1193	0.0893	0.1207	0.0797	0.0607
	Bias		0.0061	0.0018	0.0023	0.0134	0.0054	0.0009	0.0047	0.0068	0.0195

Notes: This table lists the MSE, SD and Bias of $\hat{\beta}_{X,s1}$ for simulation 1. 10'000 Monte Carlo replications have been estimated and evaluated at 9 locations in the distribution defined by the 25th, 50th and 75th percentile of both outcome variables.

Table IV: Simulation 1, MSE, SD and Bias of $\hat{\beta}_{1,s1}$

		Y_1	Q_{25}			Q_{50}			Q_{75}		
n	Crit.	Y_2	Q_{25}	Q_{50}	Q_{75}	Q_{25}	Q_{50}	Q_{75}	Q_{25}	Q_{50}	Q_{75}
400	MSE		0.5408	0.2076	0.0956	0.2810	0.1101	0.0519	0.1104	0.0441	0.0231
	SD		0.7027	0.4444	0.3056	0.5147	0.3266	0.2257	0.3251	0.2058	0.1476
	Bias		0.2168	0.1004	0.0472	0.1271	0.0582	0.0303	0.0688	0.0421	0.0368
1600	MSE		0.3459	0.2224	0.1562	0.2528	0.1650	0.1149	0.1651	0.1082	0.0800
	SD		0.3187	0.2128	0.1526	0.2413	0.1611	0.1130	0.1583	0.1025	0.0728
	Bias		0.1343	0.0647	0.0337	0.0756	0.0359	0.0207	0.0468	0.0345	0.0331
6400	MSE		0.1934	0.1192	0.0808	0.1375	0.0858	0.0594	0.0918	0.0611	0.0491
	SD		0.1562	0.1044	0.0754	0.1200	0.0788	0.0563	0.0798	0.0506	0.0366
	Bias		0.1141	0.0575	0.0288	0.0671	0.0340	0.0189	0.0454	0.0341	0.0328

Notes: This table lists the MSE, SD and Bias of $\hat{\beta}_{1,s1}$ for simulation 1. 10'000 Monte Carlo replications have been estimated and evaluated at 9 locations in the distribution defined by the 25th, 50th and 75th percentile of both outcome variables.

Table V: Simulation 2, Integrated Statistics of $\hat{F}_{Y_{1,s2},Y_{1,s2}}(t)$, different Specifications

		MDR: Cloglog NC: Gaussian PC: Normal			MDR: Logit NC: Triangular PC: Gumbel			MDR: Probit NC: Epanechnikov PC: Frank		
n	Model.	MSE	SD	Bias	MSE	SD	Bias	MSE	SD	Bias
400	MDR	0.0103	0.0394	0.0088	0.0107	0.0397	0.0091	0.0106	0.0392	0.0090
	NC	0.0166	0.0150	0.0164	0.0149	0.0254	0.0143	0.0151	0.0242	0.0145
	PC	0.0146	0.0224	0.0141	0.0169	0.0146	0.0167	0.0145	0.0240	0.0139
1600	MDR	0.0089	0.0572	0.0057	0.0093	0.0588	0.0059	0.0092	0.0579	0.0058
	NC	0.0131	0.0206	0.0127	0.0126	0.0360	0.0113	0.0127	0.0342	0.0115
	PC	0.0122	0.0331	0.0111	0.0136	0.0209	0.0132	0.0123	0.0353	0.0110
6400	MDR	0.0078	0.0665	0.0034	0.0084	0.0696	0.0035	0.0082	0.0686	0.0035
	NC	0.0106	0.0247	0.0100	0.0111	0.0424	0.0094	0.0111	0.0404	0.0094
	PC	0.0107	0.0392	0.0092	0.0113	0.0244	0.0107	0.0108	0.0418	0.0091

Notes: This table lists the average MSE, SD and Bias of $\hat{F}_{Y_{1,s2},Y_{2,s2}}(t)$ *for simulation 2. Three models have been estimated: The Multivariate distribution regression (MDR), a non-parametric copula model (NC) and a parametric copula model (PC). The estimated CDF is evaluated at a grid of* 25×25 *points defined by the quantile values of* $Y_{1,s2}$ *and* $Y_{1,s2}$*. The results of this table are based on a mean-zero normal distribution with* $\rho_{Y_{1,s2},Y_{2,s2}} = \sin(4\pi \cdot X)$ *and* $\sigma_{Y_{1,s2}} = \sigma_{Y_{2,s2}} = (X - .273)^4$ *(DGP 3 from table I). The results were obtained using 10'000 Monte Carlo replications.*

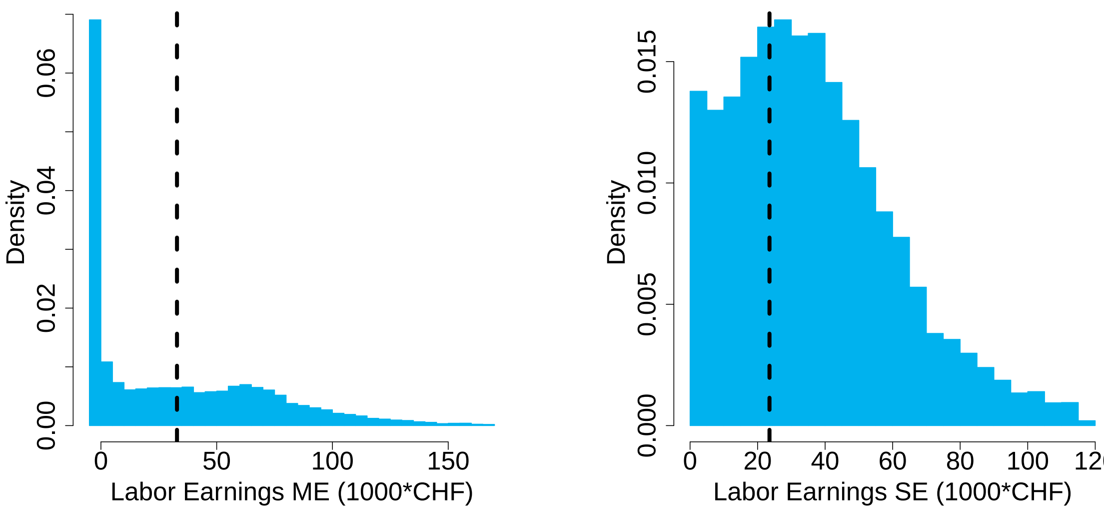
C Application Results

Table VI: Application to Household Labor Supply, Descriptive Statistics

Variable	Mean	SD	Min	$Q(.25)$	Median	$Q(.75)$	Max
Labor Income ME	35024.03	45470.60	-298193	0	20298	60990	1528772
Labor Income SE	23984.70	29886.60	-63738	0	14734	40180	495450
ME Treated	0.53	0.50	0	0	1	1	1
ME Disability Degree	39.14	40.51	0	0	42	75	100
ME Female	0.07	0.25	0	0	0	0	1
ME Age	54.39	8.03	25	50	56	61	65
ME Swiss	0.90	0.30	0	1	1	1	1
Number of Children	0.61	0.98	0	0	0	1	7
Costs: Illness	1538.60	2791.32	0	0	0	2418	81914
Costs: Healthcare	3266.67	2378.71	0	0	3500	4900	48515
Costs: Being Handicapped	222.57	2479.58	0	0	0	0	136965
Net Wealth (1000 CHF)	287.36	810.38	-3259	0	87	361	40000

Notes: This table presents the descriptive statistics for the esitmation sample. The number of observations is $N = 30,604$. Note that the 'costs' variables refer to the amount of arising costs that can be deducted from the taxable income.

Figure 6: Histogram Labor Income



(a) Main Earner

(b) Second Earner

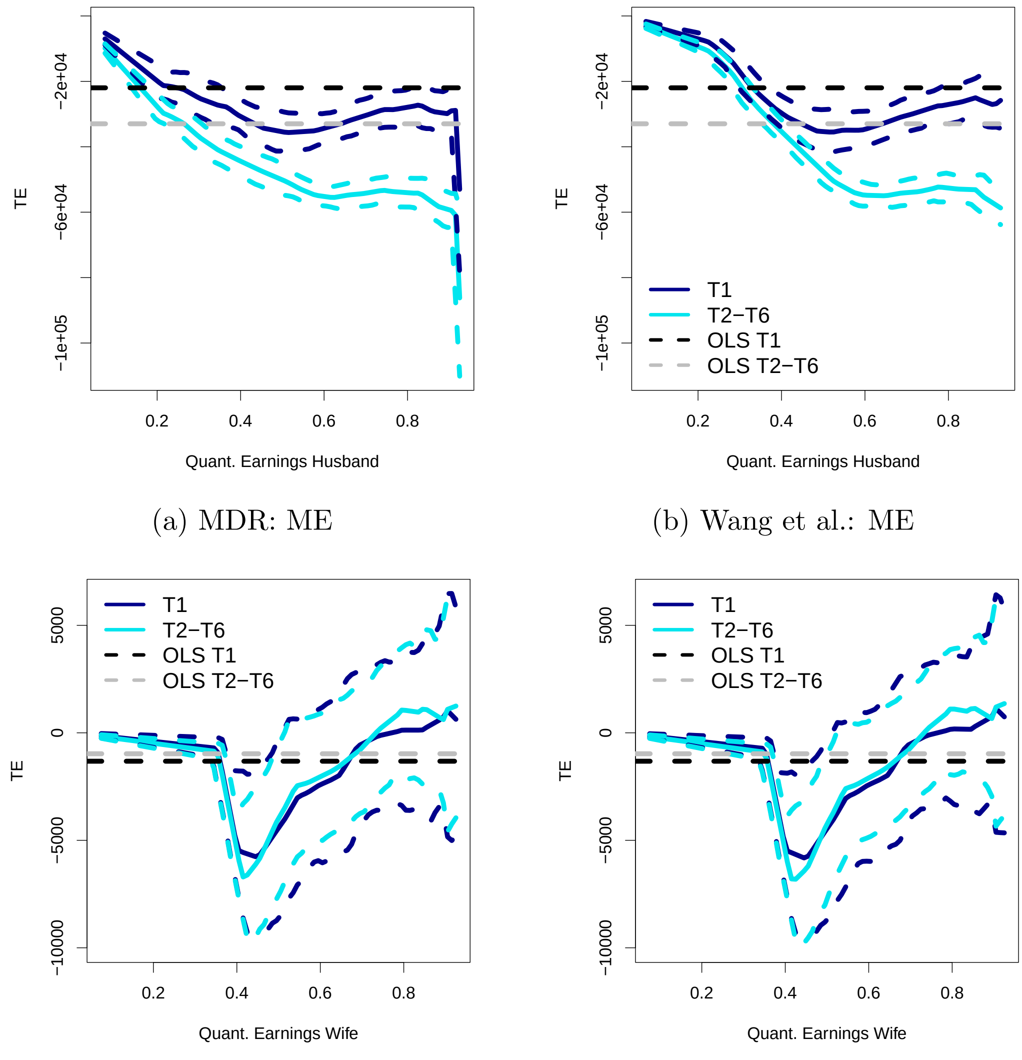
Notes: This figure presents the unconditional distributions of main and second earners labor income. The dotted black line refers to the average earnings (i.e. 90,355 CHF for MEs and 27,231 CHF for SEs). The bottom and top 1% of both distributions is not plotted to improve readability.

Table VII: Application to Household Labor Supply, Aggregate Results

	OLS				MDR			
	Earnings ME		Earnings SE		Earnings ME		Earnings SE	
	Coef	SD	Coef	SD	TE	SD	TE	SD
ME Disability Degree	-445.56	6.87	-13.16	4.95	-50930.07	-219.1	-934.49	26.7
ME First Year Treated	1165.73	1923.39	1504.45	1384.32				
ME Degree∗First Year	132.89	25.06	-25.03	18.04	-30532.21	171.41	-1297.51	166.73
ME Female	-14336.94	947.64	-23598.36	682.05				
ME Age	5331.53	293.96	1973.37	211.57				
ME Age Squared	-52.5	2.92	-21	2.1				
ME Swiss	11313.14	808.49	-1751.55	581.89				
Number of Children	2027.07	278.43	337.37	200.4				
Costs: Illness	0.04	0.09	0.17	0.06				
Costs: Healthcare	1.22	0.13	0.78	0.09				
Costs: Being Handicapped	0.12	0.09	0.09	0.07				
Net Wealth (1000 CHF)	8.74	0.29	2.99	0.21				
Region FE		Y		Y		Y		Y
Year FE		Y		Y		Y		Y
R2		0.22		0.07				
F-Statistic		400		103				
N		30604		30604		30604		30604

Notes: This table presents the OLS and MDR results for the esitmation sample. In the case of OLS, the numbers refer to the regression coefficients whereas for the MDR results, 'TE' refers to the treatment effect. The latter is defined as the average earnings implied by the difference of the two marginal, counterfactual distributions (average disability degree/no disability in the first year treated or later). The standard errors for the MDR estimator have been obtained using 250 bootstrap draws. Note that the 'costs' variables refer to the amount of arising costs that can be deducted from the taxable income.

Figure 7: Univariate Quantile Treatment Effects

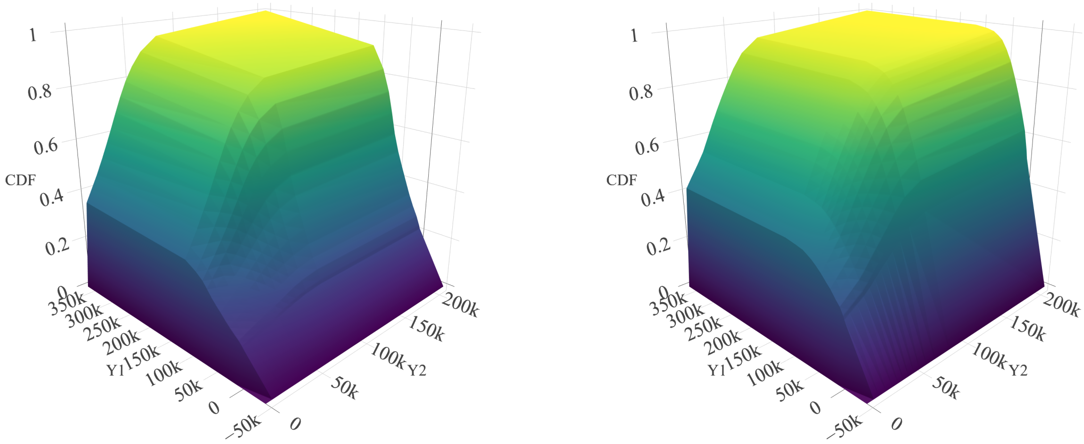


(a) MDR: ME (b) Wang et al.: ME

(c) MDR: SE (d) Wang et al.: SE

Notes: Panels (a)-(d) present univariate quantile treatment effects implied by the difference of the two counterfactual distributions (average disability degree/no disability) estimated either by MDR or the estimator of Wang et al. (2022). T1 refers to the first period of being treated and T2-T6 for the remaining periods. The dotted black and grey lines show the TE implied by an OLS regression. The dotted blue lines represent uniform 95%-confidence bands computed according to Algorithm 1 in section 3 with 250 bootstrap draws. The number of observations is 30,604.

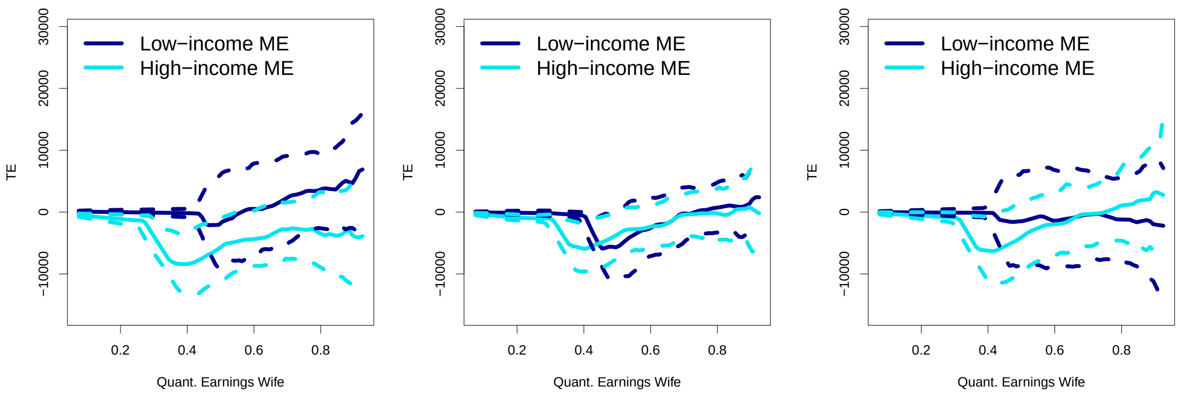
Figure 8: Bivariate Distribution of Labor Earnings



(a) Untreated (b) Treated (second to sixth year)

Notes: This figure presents the two counterfactual distribution of labor earnings outlined in section 5. Y_1 denotes the main earners labor earnings whereas Y_2 denotes the second earners labor earnings.

Figure 9: Quantile Treatment Effects: First year



(a) MDR (b) Wang et al. (2022) (c) Univariate DR

Notes: Panels (a)-(c) present quantile treatment effects implied by the difference of the two counterfactual distributions (average disability degree/no disability). According to equation equation (12) and (13), panel (a) shows QTE on the second earners earnings conditional on whether the main earner has an income below or above the median. Panel (b) presents the analogue result for the joint distribution implied by the estimator of Wang et al. (2022). Panel (c) shows the results of a univariate DR approach where a dummy variable for low/high income of the ME was included. The dotted lines represent uniform 95%-confidence bands computed according to Algorithm 1 in section 3 with 250 bootstrap draws. The number of observations is 30,604.