Three Essays in Econometrics: Heterogeneity across Outcomes, Time and Physicians

Inauguraldissertation zur Erlangung der Würde eines

Doctor rerum oeconomicarum

der Wirtschafts- und Sozialwissenschaftlichen Fakultät
der Universität Bern

vorgelegt von
Jonas Christoph Meier
von Winterthur, Zürich

April 2021

The faculty accepted this thesis on the 21. May 2021 at the request of the two reviewers Prof. Dr. Blaise Melly and Prof. Dr. Martin Huber as dissertation, without wishing to comment on the views expressed therein.

Die Fakultät hat diese Arbeit am 21. Mai 2021 auf Antrag der beiden Gutachter Prof. Dr. Blaise Melly und Prof. Dr. Martin Huber als Dissertation angenommen, ohne damit zu den darin ausgesprochenen Auffassungen Stellung nehmen zu wollen.

Preface

First and foremost, I would like to thank my supervisor Blaise Melly for his support and guidance. His advice contributed substantially to the present thesis. Both the many helpful discussions and his insightful feedback taught me a lot. A special thank goes to Martin Huber, who has agreed to act as an expert for this thesis.

Further, I owe a special thanks to my co-author Tamara Bischof. Her excellent writing and her critical thoughts improved our joint work considerably. I benefitted from her experience in organizational matters, too. For instance, this includes the faculty meetings we attended in our roles as Ph.D. representatives and many other issues. She made it much easier for me to find my way as a junior researcher.

I want to thank my faculty colleagues at the University of Bern for their valuable comments and suggestions. My thesis benefited extensively from the helpful feedback and remarks during the many Brown bag seminars or over coffee. I was lucky enough to spend many hours with my inspiring fellow Ph.D.'s. In particular, I thank Armando Näf for his patience during the endless whiteboard sessions, Nadia Ceschi for cheering me up, Severin Lenhard for his industrial assistance, Marc Brunner for his pragmatism, Fabio Canetg for his straight opinions, Nana Adrian for her moral support, Martina Pons for sharing a passion, Dino Collalti for his support on techniques, and Sarah Spycher for her enthusiasm. Also, I thank Fiona Scheidegger, Marlies Wälchli, Susanne Moser and Gabriella Winistörfer who supported me in organizational affairs and numerous other circumstances.

My final and most profound thank goes to Ramona. In many dimensions, you have been the key during the progress of this thesis. You critically reflected on my work when needed but supported me thoroughly in more challenging moments. Thank you for your time and patience, for placing back your interests, and for putting everything in perspective.

Bern, April 2021 *Jonas Meier*

Contents

Preface **v**

Contents **vi**

List of Figures **ix**

List of Tables **x**

Introduction **xi**

1 Multivariate Distribution Regression **1**
1.1 Introduction 2
1.2 Concept and Examples 5
1.3 Asymptotic Theory 8
1.3.1 Assumptions 8
1.3.2 Results 11
1.3.3 Counterfactual Distributions 13
1.3.4 Testing 15
1.4 Simulations 16
1.5 Application to Household Labor Supply 18
1.5.1 Data 20
1.5.2 Model 21
1.5.3 Results 22
1.6 Conclusion 24
1.A Theoretical Results 26
1.A.1 Proof of Theorems 1 and 2 26
1.A.2 Explicit Forms of the Link Functions 28
1.B Simulation 30
1.B.1 Misspecification 30
1.B.2 Scaling P(X) 30
1.B.3 Bandwidth Simulation 31
1.C Application Results 34

2 Censored Distribution Regression **41**
2.1 Introduction . . . 42
2.2 Related work . . . 45
2.3 Model, Identification and Estimator . . . 46
2.3.1 Redistribution of Mass to the right . . . 49
2.3.2 Martingale-based Approach . . . 51
2.4 Asymptotic Results . . . 53
2.5 Simulations . . . 56
2.6 Application to Unemployment Duration . . . 61
2.7 Conclusion . . . 66
2.A Simulation Results . . . 68
2.B Application Results . . . 71

3 Physician Induced Demand and Financial Incentives **73**
3.1 Introduction . . . 74
3.2 Institutional Background . . . 77
3.2.1 Provider Reimbursement . . . 78
3.2.2 2018 Reform . . . 78
3.2.3 Exogeneity of the Fee Changes . . . 80
3.3 Data . . . 81
3.4 Descriptives and Graphical Evidence . . . 82
3.4.1 Graphical Evidence . . . 82
3.5 Empirical Strategy . . . 85
3.5.1 Substitution Across Services . . . 86
3.5.2 Total Healthcare Supply . . . 90
3.6 Results and Discussion . . . 93
3.6.1 Substitution across Services . . . 93
3.6.2 Total Healthcare Supply . . . 96
3.6.3 Which Channel is More Important? . . . 99
3.6.4 Limitations . . . 100
3.6.5 Policy Implications . . . 101
3.7 Conclusion . . . 102
3.A Calculation of Relative Value Units . . . 104
3.B Compositional Changes . . . 106
3.C Assessing the Importance of the Substitution and the Volume Expansion Channel . . . 108
3.D Prediction of Physican-Specific Service Volumes . . . 111
3.E Additional Tables . . . 112
3.F Additional Figures . . . 118

Bibliography **123**

Statement of Authorship **137**

List of Figures

1.1 Quantile Function Second Earner 23
1.2 Bivariate Distribution of Labor Earnings 38
1.3 Histogram Labor Income 39
1.4 Distribution of DI Benefits and Quantile Effects 39

2.1 Distribution of the $\hat{\beta}_{x_u}(y)$ for CDR-MLE and CDR-Martingale 59
2.2 Bias of $\hat{\beta}_{x_u}(y)$ as a Function of Y 60
2.3 Effect of longer PBD: Coefficients 65
2.4 Effect of longer PBD: Quantile Treatment Effect 66
2.5 Outcome and Censoring Distribution for all DGP 68

3.1 Distribution of Pre-Reform Fees and Fee Changes 80
3.2 Total and Predicted Costs (Quarterly) 84
3.3 Substitution Pattern 95
3.4 Substitution Pattern by Physician Specialty 96
3.5 Total Healthcare Supply 97
3.6 Pre-Treatment Changes in Service Volumes 118
3.7 Decomposition of Changes in Revenue, 2017 to 2018 119
3.8 Total and Predicted Costs, Estimation Sample (Quarterly) . . 120
3.9 Distribution and Standard Deviation of BITE 120
3.10 Statistics for the Top 12 specialties 121
3.11 Pre-reform Consultation Growth by BITE 121

List of Tables

1.1 IMSE for Copula and DR models 18
1.2 Disability Main Earner, Average Treatment Effects 23
1.3 Parametric Link Functions 29
1.4 IMSE for Copula and DR models, No Variance Correction . . 31
1.5 IMSE for Copula and DR models under Misspecification . . 32
1.6 Scaling P(X) 32
1.7 Bandwidth Simulation, $X \sim N(0,1)$ 33
1.8 Descriptive Statistics, Individuals living in Couples 34
1.9 Differences within Couples 35
1.10 Disability Main Earner, All Statistics 35
1.11 Disability Second Earner, All Statistics 36
1.12 Treatment Effects, All Statistics, No Health Covariates 37

2.1 Average MSE, Bias and SD of the CCDF (multiplied by 100) . 58
2.2 Corrected Average MSE, Bias and SD of the CCDF (multiplied by 100) 69
2.3 Continuous Case ($T = 100$): Average MSE, Bias and SD of the CCDF (multiplied by 100) 70
2.4 Descriptive Statistics by Treatment and Control Group 71

3.1 Descriptive Statistics, Quarterly Means 83
3.2 Descriptive Statistics: Quarterly BITE by Medical Specialty (Top 12) 112
3.3 Descriptive Statistics: BITE by Medical Specialty (Non-Top 12) 113
3.4 Substitution Model: Regression Estimates 114
3.5 Substitution Model: Regression Estimates II 115
3.6 Total Healthcare Supply: Regression Estimates (Consultations)116
3.7 Total Health Care Supply: Regression Estimates (Revenue Decomposition) 117

Introduction

This thesis consists of three essays in econometrics. The first part of the thesis, consisting of chapters one and two, treats novel estimators aiming at distributional effects. In the first chapter, I introduce a methodology to estimate the joint distribution of multiple outcome variables. The second chapter incorporates censoring, a prevalent challenge when analyzing duration data, into distribution regression methods. Finally, the third chapter analyzes a reform in the health care sector.

Chapter 1 introduces Multivariate Distribution Regression (MDR), a semi-parametric approach to model the joint distribution of several outcome variables. Typically, researchers are interested in the effects on multiple outcomes when the latter are correlated (see Patton, 2012, for an overview). For instance, this is the case for the impact of a crisis on asset prices. Asset prices co-move tightly as they depend on common factors such as market cycles. A shock on one price index may thus affect many other indices. In addition, the effect could vary across the distribution of the prices - a peculiarity that MDR accounts for. Essentially, MDR estimates the impact of interest at every point of the outcome's distribution.

MDR's most obvious advantage is its flexibility. Existing methods, such as copula models, typically impose a parametric form of the dependence structure across outcomes (i.e. Klein et al., 2019). In contrast, MDR does not require equally restrictive, parametric assumptions. Thus, the effects estimated using MDR describe the underlying mechanisms more accurately. Further, MDR generalizes two well-known estimators: (i) the empirical multivariate cumulative CDF by allowing for covariates and (ii) univariate Distribution Regression (DR) by considering multiple outcomes. Building on earlier work in the field (Chernozhukov et al., 2013), I establish that MDR consistently estimates the regression coefficient process. Further, I show that coefficients are well-behaved and converge to a Gaussian process, with the bootstrap being a consistent tool to assess the asymptotic distribution.

To illustrate the usefulness of MDR, I estimate the effect of disability insurance benefits on labor supply responses among Swiss households. Generally, receiving these benefits is related to lower incentives to supply labor (i.e. Autor et al., 2016). Autor et al. (2019) find that spouses increase their labor supply once their partner is disabled. My results indicate that spouses of low-income partners do respond as suggested by Autor et al. (2019). Yet, among average to high-income households, the need to compensate for the financial loss appears less immediate.

In Chapter 2, co-authored with Blaise Melly, we incorporate censoring into the univariate DR model. The resulting estimator, censored distribution regression (CDR), allows studying how the covariates' effects vary over time. From a theoretical perspective, CDR represents a generalization of three existing estimators. In particular, CDR simplifies (i) to the Kaplan-Meier estimator in the absence of covariates Kaplan and Meier (1958), (ii) to distribution regression in the absence of censoring, and (iii) to Cox's proportional hazard estimator in the absence of heterogeneity (Cox, 1972). As our main results, we establish weak convergence of the coefficient process to a Gaussian process.

The standard tool to analyze duration data is Cox's proportional hazard model, which assumes time-constant effects. On many occasions, this assumption seems too restrictive. For instance, job search behavior differs during unemployment. In this context, we apply the CDR estimator to estimate the effect of potential benefit duration (PBD) on unemployment spells. Search models suggest that faced with the upcoming exhaustion of benefits, individuals intensify their search efforts and lower their target wages (Krueger and Mueller, 2016; Marinescu and Skandalis, 2021). Our results indicate that PBD has a negligible effect for short-term unemployed but a strong and significant effect for the long-term unemployed. This is in line with an increased likelihood of finding a job once the benefits are close to exhaustion.

In Chapter 3, co-authored with Tamara Bischof, we address how physicians respond to changes in their financial incentives. We exploit plausibly exogenous changes in the fee structure for medical services in the outpatient sector. The tariff partners, the health care providers and insurances, failed to reach an agreement on how to reform the outdated tariff scheme *TARMED*. In response, the federal government set the new fees, causing a revenue loss of up to 40% for single physicians. Previous research suggests that physicians may respond in two different ways: Faced with a revenue loss, physicians can (i) substitute from low-paying to more attractive services and (ii) increase their overall health care supply (i.e. Clemens and

Gottlieb, 2014; McGuire and Pauly, 1991; Yip, 1998). Our main goal is to disentangle these two channels and to quantify their relative importance.

Our results are threefold. (i) We find that providers raise (lower) the volume of services that have become relatively more (less) attractive. (ii) Physicians increase their overall volume of services and treat more patients when they lose a significant share of their revenue. (iii) Finally, a comparative exercise indicates that volume expansions are far more important than substitution responses. In particular, a revenue loss of 5% leads to an increase in the overall supply of roughly 3% whereas we do not observe a significant rise in substitution responses. Concerning policy implications, our results suggest (i) that gradual fee changes may prevent strong and costly reactions due to more considerable revenue losses. (ii) Further, policy-makers could directly incentivize physicians to provide services that are of high value for consumers.

Chapter 1

Multivariate Distribution Regression

Abstract

This paper introduces multivariate distribution regression (MDR), a semi-parametric approach to estimate the joint distribution of outcomes. 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 CDF. For instance, this includes counterfactual distributions. Compared to copula models, MDR achieves the same accuracy but is (i) more robust to misspecification and (ii) allows to condition on many co-variates, thus ensuring a high degree of flexibility. Finally, an application analyzes 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. Whereas the opposite

Acknowledgements: I would like to thank Blaise Melly for valuable comments and suggestions. I would also like to thank participants at the Brown Bag Seminar at the University of Bern for instructive comments. 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 would like to acknowledge the financial support from the Swiss National Science Foundation for the grant 165621.

holds for high-income households, likely because they are well insured and can afford to work fewer hours.

1.1 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 important 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 are 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. Aloui et al., 2011; Christoffersen et al., 2012). 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) generalizes two well-studied estimators: (i) the empirical multivariate CDF by including covariates and (ii) univariate Distribution Regression (DR) by considering multiple outcomes. In its simplest form, MDR only includes a constant as a regressor. In this case, MDR reduces to the empirical multivariate CDF as the CDF is approximated separately in each cell. In contrast to univariate DR, MDR provides information on the dependence structure, i.e. 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 contributes by offering many novel possibilities to analyze the joint CDF. For instance, this includes quantile functions of one outcome conditional on others, unconditional distributions, and counterfactual distributions. In the following, I sketch the concept of the latter to highlight the additional value for applied research.

In the context of MDR, a counterfactual is a CDF modified by one of two types of hypothetical changes: (i) a change in the distribution of covariates or (ii) the conditional distribution itself. To illustrate the former, suppose the researcher is interested in the effect of a binary treatment. Intuitively, the counterfactuals answers to questions as "How would the CDF look like if all (no) observations were treated?". Comparing the CDFs then identifies the distributional treatment effect, provided that the treatment is randomly assigned. Note that the distributional treatment effects may vary with the level of the outcomes, which introduces a natural way to address heterogeneity. The second type of changes refers to questions as "How would the CDF for treated units look like if the covariates had the same effects as for untreated units?". Asking such questions enables comparisons between subgroups of the data.

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 one member newly receives 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 subsidised, this effect has drawn a lot of 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. MDR enables to analyze these shifts separately at all parts of the bivariate distribution of labor earnings. I find that the spouses' response crucially depends on the income of their partner. If low-income main earners are hit by the shock, 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 take care of their partners and the household chores as they can afford to work fewer hours.

The estimator builds on the theory of univariate distribution regression 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 with DR. This paper departs from the setting in Chernozhukov et al. (2013). Yet, I take a different approach when proving consistency and asymptotic normality.[2] Related developments in the field include the possibility to in-

[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).

[2]Using theorems for approximate Z-estimators, Chernozhukov et al. (2013) derive the asymptotic distributions for functionals of quantile and distribution regression based

troduce two-way fixed effects into DR models (Chernozhukov et al., 2020a) or the estimation of structural functions in nonseparable triangular models (Chernozhukov et al., 2020b).

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 non-trivial 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 favour 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). Focusing on conditional processes, recent studies tackled the issue using copula models (Fermanian and Lopez, 2018; Portier and Segers, 2018). In general, copulas are commonly used to model multivariate CDFs (e.g. Patton, 2012). Copula models are attractive due to the possibility to separately specify the marginals and the dependence structure, the copula itself, ensuring a high degree of flexibility. Yet, the assumption on the copula is crucial and may be too restrictive (Ho et al., 2015; Zimmer, 2012). Possibly, the issue may be solved by nonparametric estimates of the copula (Gijbels et al., 2011). However, as the number of regressors increases, these models are infeasible in practice because they suffer from the curse of dimensionality (Fermanian and Lopez, 2018). Similarly, the non-parametric approach of Bouzebda and Nemouchi (2019) using U-processes is likely to suffer from the same drawback. 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. In contrast to

---

estimators simultaneously. The present proofs are based on theorems regarding exact Z-estimators as quantile regression methods play no part in this analysis.

MDR, this model does not simplify to empirical multivariate CDF in the absence of covariates. Thus, the approach imposes additional assumptions on the joint distribution. Further, MDR is more general in the sense that the choice of the link function, the analogue to the transformation function, is less likely to affect the results. In a simplistic simulation setting, I show that MDR outperforms copula models even in cases with only one regressor.

The remainder of this paper is structured as follows. Section 1.2 introduces the model and outlines two typical settings where MDR and counterfactual CDFs may be advantageous. Section 1.3 derives the asymptotic theory of the estimator. A simulation example is set up in section 1.4. Section 1.5 presents the application and section 1.6 concludes.

1.2 Concept and Examples

This section introduces the approach of MDR by first presenting the model of the multivariate CDF and then focusing on desired functionals of the CDF, primarily counterfactual distributions. At the end of this section, I will sketch two examples to illustrate the wide range of potential applications.

At the core of the model, MDR estimates the joint CDF of multiple outcomes. Considering a bivariate case, denote the outcomes by Y_1 and Y_2. The joint CDF is the probability of Y_1 and Y_2 being smaller than some specified values, say t_1 and t_2. Formally, this is $F_Y(t) = P(Y_1 \leq t_1, Y_2 \leq t_2)$, where t is the vector of thresholds. Equivalently, $F_Y(t)$ is the expected value of the binary variable $E[\mathbf{1}(Y_1 \leq t_1, Y_2 \leq t_2)]$. Most applied research is interested in how regressors affect this probability. A natural possibility to model $E[\mathbf{1}(Y_1 \leq t_1, Y_2 \leq t_2)|X]$ 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 $F_{Y|X}(t|X)$. Essentially, the MDR estimator models this conditional expectation at a large number of thresholds. The resulting coefficients on X_i, $\beta(t)$, are allowed to vary with t which ensures a high degree of flexibility. This modelling approach entails several attractive features. First, each individual regression is tractable and offers various forms of well-known results such as marginal effects. Second, the estimator is trivial to implement. Third, the obtained fit of the CDF can be used to estimate any feature of the joint distribution. Among others, this includes averages and variance-covariance matrices. Further, taking all outcomes but one to the limit, one can derive the marginal CDF which, in turn, implies the marginal PDF. The quantile function can be obtained by taking the left inverse of the marginal CDF. Taking derivatives, the multivariate PDF is implied, too.[3] Note that the PDF is consistently estimated but converges at a lower rate than the distribution itself (Rothe and Wied, 2020). Finally, it is feasible to derive the marginal CDF conditioning on values of the other outcomes.[4] Taken together, the multivariate CDF fully describes the statistical properties of the outcomes.

Prior to addressing potential applications, I show how the estimated CDF can be used to answer specific questions. For this purpose, I will focus on counterfactual distributions as an example of the functionals that are implied by the joint CDF. These distributions describe how the CDF depends on two types of hypothetical changes. On one hand, changes in *covariates* could result in a new shape of the CDF. On the other hand, the *effects* of the covariates may be altered, this is changing the conditional distribution itself. To illustrate the former case, suppose we are interested in the effect of a binary variable D_i which is included in 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 provides an answer to "How would the CDF look like if everyone (no one) was treated by the effect of D_i?".[5] The comparison of these two CDFs has a causal interpretation, provided that the treatment is randomly assigned once conditioned on the covariates. Similarly, we can target any unit change of a discrete or continuous variable in X_i. Extending this idea, the whole distribution of covariates may be altered. More specifically, in a model for men and women wages, questions like "How would the CDF look like if women had the characteristics of men?" could be answered. The second type of modifications is concerned with the CDF for different subgroups too. Again referring to men and women, it could be analyzed how women's outcome distribution would look like if the covariates had the same *effects* on wages as for men. Formally, this consists in changing $F_{Y|X}(t|X)$. Finally, note that both types of modifications may be considered simultaneously. Next, the application of MDR is sketched in two typical examples.

Example 1. Bivariate Labor Supply Family labor supply has gained a lot of attention due to its relevance for intra-family and aggregate inequality (Eika et al., 2019; Hyslop, 2001; Schwartz, 2010). Consider two spouses who both

[3] These approximations rely on numerical derivatives. Thus, it is crucial that a fine enough grid of $t \times t$ thresholds has been used to estimate the model.

[4] In the bivariate case, the marginal CDF of Y_2 at the median of Y_1, $Q_{50}(Y_1)$, is $F_{Y_2|Y_1=Q_{50}(Y_1)}(t) = \frac{F_Y(t)(Q_{50}(Y_1),Y_2)}{F_Y(t)(Q_{50}(Y_1),\infty)}$.

[5] Formally, these CDF's are $F_{Y\langle\cdot|D=0\rangle}(t) = \int_{\mathcal{X}_i} F_{Y|X_i}(t|X)dF_{X_{i,D_i=0}}(X)$ and $F_{Y\langle\cdot|D=1\rangle}(t) = \int_{\mathcal{X}_i} F_{Y|X_i}(t|X)dF_{X_{i,D_i=1}}(X)$ where $\cdot$ indicates that the conditional distribution did not change.

participate in the labor market. Let Y_1 and Y_2 denote labor income of the woman and man, respectively. Notably, labor income depends on a number of factors such as gender, age, education, parenthood, origin or industry (see e.g. Blau and Winkler, 2018, for an overview). Further, in response to persistent wage shocks, family labor supply may act as insurance (Blundell et al., 2016) and the response depends on the complementarity of leisure time and the substitutability of childcare (Blundell et al., 2018). In fact, the presence of children crucially determines the spousal labor supply response (Halla et al., 2020). Thus, spouses' bivariate distribution of labor supply is likely to be a function of household characteristics. The latter should be included as regressors to avoid misspecification.

Researches may be interested in the effect of women's level of education on the joint distribution of earnings. Using counterfactuals, the CDF could be modelled for every level of education. For instance, suppose we only care about the correlation of male and female's labor income. Comparing the counterfactual distributions, the correlation could be viewed as a function of women's level of education. As shown by Goussé et al. (2017), education induces assortative mating.[6] This finding can be tested by treating a positive correlation of spousal income as a sign of assortative mating. MDR could extend this analysis by considering more complex measures such as tail dependences.

One could aim at estimating the causal effect of a binary income shock on the bivariate distribution of labor income. For instance, a tax reform or winning the lottery may serve as treatment variables. Assuming that the treatment is randomly assigned conditional on the included covariates, the counterfactual distributions in the presence (absence) of the shock have a causal interpretation.[7] Cesarini et al. (2017) find that winning the lottery modestly reduces labor supply for winners. The reduction is smaller for spouses, which is inconsistent with unitary household models. Using MDR, modelling 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. In the context of tax reforms, it is of interest how households would react to a change from joint to individual taxation for married couples. This is currently discussed in Switzerland.[8]

[6]Note that in their work, Goussé et al. (2017) analyze exogenous variations in education.

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

[8]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). Yet, the political parties have not reached an agreement (CH Media,

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, job loss or retirement.

Example 2. Asset Prices The interdependence of asset prices in multiple markets has long been recognized (e.g. Aloui et al., 2011; Chavas, 2020; Christoffersen et al., 2012). In short, positive correlations of asset prices arise across countries and markets. These correlations depend on numerous factors such as market cycles and the type of the goods and the stage of development of the economy. The resulting linkages are crucial for returns as they determine diversification benefits and portfolio allocations. Typically, empirical studies estimate the joint distribution of equity price indices of numerous countries (Christoffersen et al., 2012).

In the following, consider price indices of different countries as outcome variables. Further, suppose that the regressors include market and country variables as well as specifics of the underlying goods. In this context, the MDR estimator provides two appealing features. First, we can account for the apparent nonlinear dependencies and the asymmetric tails by including a potentially large number of regressors. In addition, the coefficients are allowed to vary across the distribution of prices, which introduces further flexibility and ensures that the CDF is well approximated. Second, considering treatment effects of price shocks, counterfactual correlation matrices could be defined. This abstract exercise could compare the CDF and its implied correlations in the presence or absence of an economic shock. Thus, correlation patterns may differ depending on whether the shock affects the industry, the country or global markets. Examples of shocks on asset prices include the great recession, monetary policy decisions or non-financial shocks as the COVID-19 pandemic. The MDR estimator may contribute to the literature on the great recession as it improves upon copula approaches since the latter are likely to be misspecified (Zimmer, 2012). Alessi and Kerssenfischer (2019) show that large-scale dynamic factor models are able to indentify monetary policy shocks, once the regression includes many observables. As MDR allows for the inclusion of large number of covariates, the model fits well into this framework and represents an alternative to existing approaches. Finally, Caballero and Simsek (2020) find that COVID-19 reduced asset prices through different channels. In this setting, MDR could show how the raising pandemic in one region affects the price indices

2020).

of another region. Rodriguez (2007) analyzed financial contagion using copula models. Replicating the same study using MDR may contribute by addressing whether the parametric assumptions on the copula affect the results.

1.3 Asymptotic Theory

1.3.1 Assumptions

This section introduces and discusses the assumptions on the underlying data and the model. 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 and X_i is a set of K regressors.

Assumption 1. Data *The data Y_i, X_i is i.i.d with bounded support.*

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

$$F_{Y_i|X_i}(t|X_i) = \Lambda(P(X_i)'\beta(t)), \tag{1.3.1}$$

where $\beta(t)$ is a $K \times 1$ function-valued coefficient vector and $P(X_i)$ is a $1 \times K_p$ 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 $P(X_i)$ is rich enough, the CDF is approximated arbitrarily well and the parametric form of the link function is irrelevant too.[9] Additional remarks should be made with respect to assumption 2. First, the estimator for $\beta_0(t)$ defined in equation (1.3.1) is

[9] Following the argument in Chernozhukov et al. (2013, p. 2217), let $P(X)$ consist of the first p components of a basis in $L^2(\mathcal{X}, P)$. Assume that $\Lambda^{-1}\left[F_{Y_i|X_i}(t|X)\right] \in L^2(\mathcal{X}, P)$ and $\lambda(z) = \frac{\partial\Lambda(z)}{\partial z}$ is bounded above by $\bar{\lambda}$. Define the squared misspecification

a Z-estimator and can be interpreted as a pseudo maximum likelihood estimator because it provides the best approximation of the CDF given a specific link function. Second, the model in equation (1.3.1) 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, $P(\cdot)$ indicates that the model allows for the inclusion of a wide range of functionals of X_i. In the remainder of the paper, $P(X_i)$ is denoted by X_i with dimension $N \times K$ for simplicity. Finally, the distribution function equation (1.3.1) does not to be continuous and the model also captures the discrete outcome variables.[10]

The objective function is defined as an approximate zero of a function $\Psi(\beta, t)$ between two normed spaces. More precisely, $\Psi(\beta, t) : \mathcal{M} \times \Theta \mapsto \Theta$ where $\mathcal{M}$ is an open set containing $\mathcal{T}$ and $\Theta = \mathbb{R}^{d_K}$ is the parameter space which contains β. 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

$$\psi_{\beta,t} = \left(\Lambda\left[X_i'\beta(t)\right] - y_i(t)\right)\left(\frac{\lambda\left[X_i'\beta(t)\right] X_i}{\Lambda\left[X_i'\beta(t)\right]\left(1 - \Lambda\left[X_i'\beta(t)\right]\right)}\right), \quad (1.3.2)$$

where $\lambda(\cdot)$ denotes the derivative of $\Lambda(\cdot)$. Accordingly, let $\Psi_n(\beta, t) = \mathbb{P}_n\psi_{\beta,t}$ be the sample estimator, where $\mathbb{P}_n$ is the empirical measure. The subsequent assumption is concerned with the properties of $\Psi(\beta, t)$ itself. Then, assumption 4 introduces the requirements for the bootstrap to be a valid tool to do inference.

Assumption 3.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 4.Bootstrap *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 $\mathbb{P}_n^\circ f =$*

---

error $\delta_p = E\left[\Lambda^{-1}(F_{Y_i|X_i}(t|X)) - P(X)'\beta(t)\right] E\left[\Lambda^{-1}(F_{Y_i|X_i}(t|X)) - P(X)'\beta(t)\right]'$. Then, it can be shown that $\delta_p \to 0$ as p grows. Thus, $E\left[F_{Y_i|X_i}(t|X) - \Lambda(P(X)'\beta(t))\right] E\left[F_{Y_i|X_i}(t|X) - \Lambda(P(X)'\beta(t))\right]' \leq \bar{\lambda}\delta_p \to 0$ by weak concavity of Λ.

[10]To see this, suppose that the true CDF $F_{Y_i|X_i}(t|X_i)$ is discrete or mixed. As the approximation is done 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 (1.3.1) 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.

$n^{-1}\sum_{i=1}^{n}\frac{\xi_i}{\bar{\xi}}f(X_i)$ 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 3 requires the objective function to be zero only at the true values $\beta_0(t)$. Thus, assumption 3 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 3 is seen as a technical requirement which is likely to be met.

With respect to assumption 4, 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 by the means of maximum likelihood estimation (MLE). To see this, note that the model in equation (1.3.1) is estimated as single regressions at every entry of t. For a given t, $y_i(t) = \mathbf{1}(Y_{1,i} \leq t_1, ..., Y_{d,i} \leq t_d)$ is regressed on X_i, for instance using a probit model. Thus, at a finite number of t, maximum likelihood estimation 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 3 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 1.3.3 then builds on this result and formally introduces the counterfactual framework.

1.3.2 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 1.A.

Theorem 1.Asymptotic Distribution *Let Assumptions 1 to 3 hold. Then, the MDR estimator* $\hat{\beta}_n(t)$ *of* $\beta_0(t)$ *in equation* (1.3.1) *satisfies* $\sqrt{n}(\hat{\beta}_n(t) - \beta_0(t)) \rightsquigarrow \dot{\Psi}^{-1}_{\beta_0,t} Z$*, where* $Z \in \ell^\infty(\mathcal{T})$ *is the tight, mean zero Gaussian limiting distribution of* $\sqrt{n}(\Psi_n - \Psi)(\beta_0, t)$*.*

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* (1.3.1)*. 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 is 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 4 hold. Denote the multiplier or multiplier bootstrapped estimator of $\hat{\beta}_n$ by $\hat{\beta}_n^{\circ}$. Then, the MDR estimator $\hat{\beta}_n(t)$ of $\beta_0(t)$ in equation* (1.3.1) *satisfies $\sqrt{n}(\hat{\beta}_n^{\circ} - \hat{\beta}_n) \rightsquigarrow k_0 Z$, where $Z \in \ell^{\infty}(\mathcal{T})$ is the tight, mean zero Gaussian limiting distribution of $\sqrt{n}(\Psi_n - \Psi)(\beta_0, t)$ and $k_0 = 1$ for the multinomial bootstrap and $k_0 = \frac{\tau}{\mu}$ in the case of the multiplier bootstrap.*

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

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).

1.3.3 Counterfactual Distributions

This section formalizes the framework outlined in section 1.2 and builds on the previously derived results. 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 (1.3.1), yet, they depend on the link function which complicates the interpretation. Instead, counterfactual distributions directly connect to potential changes of 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 (1.3.1), 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

$$F_{Y\langle I|J\rangle}(t) = \int_{\mathcal{X}_J} F_{Y_I|X_I}(t|X) dF_{X_J}(X). \tag{1.3.3}$$

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 (1.3.3) is integrated over all observation. Similarly, an empirical approach could focus on the causal effect of a binary treatment D_i where $X_{i,D_i=0}$ and $X_{i,D_i=1}$ represent the accordingly modified covariate matrices. Using equation (1.3.3), the CDF $F_{Y\langle \cdot|D=0\rangle}(t) = \int_{\mathcal{X}} F_{Y|X}(t|X) dF_{X_{i,D_i=0}}(X)$ can be interpreted as the distribution conditional on nobody being treated. Note that the symbol $\cdot$ refers to the fact that the conditional distribution has not been altered. Conditional on X_i, suppose that D is randomly assigned.[11] Then, the difference of the counterfactual CDF's, $F_{Y\langle \cdot|D=0\rangle}(t) - F_{Y\langle \cdot|D=1\rangle}(t)$, has a causal interpretation.

Next, consider changing the covariate 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 (1.3.1). Intuitively, this answers "How would

[11] Formally, this is $D_i|X_i \perp A$ where $\perp$ denotes independence and A is an indicator for being assigned to the treatment.

the CDF of subgroup J look like if the regressors had the same effects as for subgroup J?" In every 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 and X_i has bounded support, then $\int_{\mathcal{X}_J} \Lambda[\cdot]\, dF_{X_J}(X)$ satisfies the requirements of Corollary 1. The condition on the support of X_i is met by assumption 1. Further, Hadamard-differentiability of the link functions in assumption 2 is shown to hold in the Proof of Theorem 1.[12] Thus, counterfactual distributions are consistently estimated by the sample analogue of equation (1.3.3) and the bootstrap is a valid tool to do inference.

1.3.4 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 amounts to testing the null hypothesis $H_0 : \hat{\beta}_j(t) = \hat{\beta}_j(Q_{50}(Y))$ against $H_1 : \hat{\beta}_j(t) \neq \hat{\beta}_j(Q_{50}(Y))$, 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 most cases, they 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), these 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

[12]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.

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 (1.3.4). Note that the bootstrapped statistic in equation (1.3.5) is recentered at the average values of the point estimates.

$$T(t) = \frac{\sqrt{N}\,|\hat{\beta}_j(t) - \hat{\beta}_j(Q_{50}(Y))|}{s.e.(\hat{\beta}_j(t) - \hat{\beta}_j(Q_{50}(Y))} \tag{1.3.4}$$

$$T_b(t) = \frac{\sqrt{M}\,|\hat{\beta}_{b,j}(t) - \hat{\beta}_{b,j}(Q_{50}(Y)) - (\hat{\beta}_j(t) - \hat{\beta}_j(Q_{50}(Y)))|}{s.e.(\hat{\beta}_{b,j}(t) - \hat{\beta}_{b,j}(Q_{50}(Y))}, \tag{1.3.5}$$

where N denotes the number of observations M is the number of observations for each draw from the bootstrap. Note that M does not need to deviate from N. 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.

1.4 Simulations

The simulation study compares the MDR estimator to comparable copula models. For this, it suffices to take the simplest case of a bivariate distribution of outcomes where only the covariance depends on an observable regressor. Denote the outcomes by Y_1 and Y_2 and the regressor by X. I will assume the outcomes to be standard normally distributed with the covariance $2 \cdot \Phi(X'\beta) - 1$. Without loss of generality, I will assume $\beta = 2$.[13] Note that in general the distribution of X is unrestricted. In the following, I will consider the case of $X \sim N(0,1)$ as this produces approximately uniformly distributed covariances.[14] Formally, the distribution of the outcomes is

$$Y_S \sim N\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1 & 2\cdot\Phi(2x)-1\\ 2\cdot\Phi(2x)-1 & 1\end{pmatrix}\right).$$

The resulting process will be estimated in R (R Core Team, 2020) using 10'000 replications and three different data sizes with 100, 1'000 and 10'000 observations respectively. Including only a constant, the MDR estimator should provide similar results to a unconditional copula model. Thus, this

[13] Only $\beta = 0$ would affect the results as then X would be irrelevant.

[14] In Appendix 1.B, I consider a case where the correlations have higher probability to lie at -1 or 1. Then, MDR is clearly outperforming the copula models.

section compares the performances of a correctly specified copula model and the simplified MDR estimator. Moving on towards conditional models, a nonparametric as well as a parametric copula model are estimated and compared to the MDR estimator, all using just X and an intercept as regressors. Both copula models are specified in Derumigny and Fermanian (2017) and implemented in the R package `CondCopulas` (Derumigny, 2020).[15] For the nonparametric estimation, a Gaussian kernel is used and the bandwidth is chosen to be .35. For the conditional copula model, the corresponding value is .175.[16] The performances are evaluated using an integrated mean square error (IMSE) criterion which is

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

where C is the number of cells which is used to estimate the CDF, N is the number of observations in a cell j, $\hat{F}_{Y|X}(t|X)_{i,j}$ is estimated CDF of Y for one observation at threshold t given X in cell j and $F_{Y|X}(t|X)_j$ is the average true value of the corresponding CDF in cell j. At threshold t, the CDF is a Bernoulli ranodon variable with variance $p(1-p)$, i.e. $F_{Y|X}(t|X)_j(1 - F_{Y|X}(t|X)_j)$. I divide the MSE by the variance to rescale the deviations at the limits of the CDF. At these limits, the estimated CDF will be close to zero or one regardless which estimator is chosen. Thus, dividing by a large variance at the limits I scale the errors to be comparable.[17] 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. In the following, all copula models are correctly specified by assuming a normal copula. In contrast, the MDR estimator does not require such an assumption. Appendix 1.B.1 provides simulation results where the copulas are misspecified. Further, the accuracy of the MDR estimator can improved by considering a richer $P(X)$. The corresponding results are included in Appendix 1.B.2.[18] For the

[15]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.

[16]Note that there are no theoretical justification for the choice of the bandwidths. In the current setting, a simulation exercise which can be found in Appendix 1.B.3 suggests that for most bandwidths, the accuracy is about constant.

[17]Dividing by the variance of the estimator amounts to be an advantage for copula models. The distributional assumption of the copula models improves the accuracy at the tails which is where the mentioned division matters most. Table 1.4 in Appendix 1.B provides the results of Table 1.1 without dividing by the variance.

[18]Note that to capture the advantage of a rich set of covariates, the data generating processes Y_1 and Y_2 are too simplistic. Still, there is an improvement if the data size is

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.[19] The results of the simulation study are presented in Table 1.1.

Table 1.1: IMSE for Copula and DR models

Obs.	Statistic	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
100	IMSE	323.3603	267.5490	1505.5859	104.0488	115.8119
	IVAR	0.0006	0.0011	0.0388	0.0012	0.0083
	IBIAS	323.3597	267.5479	1505.5471	104.0476	115.8037
1000	IMSE	9.9369	7.8331	0.4628	0.5062	0.2353
	IVAR	0.0009	0.0040	0.0179	0.0156	0.0151
	IBIAS	9.9361	7.8291	0.4449	0.4906	0.2203
10000	IMSE	0.0433	0.0530	0.0318	0.0297	0.0286
	IVAR	0.0010	0.0025	0.0273	0.0292	0.0194
	IBIAS	0.0422	0.0505	0.0045	0.0005	0.0092

Notes: The values represent averages over 10'000 replications. (1): Unconditional Normal Copula Model, (2): MDR including only a constant, (3): Nonparametric Copula Model, (4): Parametric Normal Copula Model, (5): MDR. The statistics are averages over the 25×25 cells.

Naturally, the average $IMSE$ is higher for the unconditional models in columns (1) and (2) in Table 1.1 as they disregard the importance of X. Nevertheless, the MDR estimator in (2) slightly outperforms the correctly specified and unconditional copula model. The assumption on the copula only pays off with a large dataset, where the copula model performs better. For samples up to 1'000 observations, the DR model in column (5) is clearly outperforming the other conditional models in (3) and (4). In the limit, all three models produce an equal fit to the data. However, Table 1.5 in Appendix 1.B shows that under misspecification, the IMSE increases for the copula models up to a factor 12. For all models, the $IMSE$ is decreasing with the size of the data. Mainly, because the average bias over the cells decreases as the estimation is more precise. In contrast, the average variance has no clear trend. Overall, Table 1.1 indicates that in small samples the MDR models outperform the copula models. Considering large data, the choice of the model is less important.

fairly large.

[19]Note that the performance only marginally improves by considering a finer grid compared to this specification. Due to the distributional assumptions, the copula model do not require a comparably fine grid as the MDR estimator.

1.5 Application to Household Labor Supply

The division of labor within households has been studied for a long time (e.g. Bianchi et al., 2000; Fuwa, 2004; Shelton and John, 1996) 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 other factors. I will consider the labor supply adjustments of households in response to one spouse newly receiving DIB. Due to this shock, the incentives to supply labor change in opposite ways for the spouses. Due to disability, individuals may no longer be able to work full-time, which reduces working hours. Generally, receiving DI payments reduces the incentives to earn income for the handicapped (Autor et al., 2016; Leisibach et al., 2018; Marie and Castello, 2012). In the same spirit, losing disability insurance (DI) eligibility and the corresponding payments increase household earnings (Deshpande, 2016). However, Autor et al. (2019) find no reduction in income for the disabled but a significant increase in spousal labor supply. Likely, the latter response intends to secure the economic stability of the household. Lee (2020) finds substantially lower labor supply responses, mainly because spouses spend many hours taking care of their disabled partners. Note that I do not intend to comprehensively analyze the socio-economic circumstances driving labor supply; the application rather analyzes a shift in the division of labor.

I argue that MDR identifies the causal effect of newly receiving DIB for three reasons. First, I control for potential confounders by including age, gender, education, health indicators and children as regressors into the model. The amount of DIB is correlated with these socio-demographic factors. Thus, including these variables is crucial for the identification of the treatment effect. In a similar setting, French and Song (2014) find that once covariates are included, the difference between ordinary least square and instrumental variable estimates is negligible. Second, it rarely happens that both partners are hit by the shock.[20] This ensures that the treatment effect is not jeopardized by a common shock on both spouses. Third, institutional features introduce considerable randomness into the process determining who will receive payments. A law passed in 2012 assigns the claims randomly to the authorized institutions, similar to the Norwegian setting (Autor et al., 2019; Dahl et al., 2014).[21] Generally, the physicians granting

[20]In particular, receiving DIB is uncorrelated within households (correlation of 0.057).

[21]The Article 72 of the IVV (*Verordnung über die Invalidenversicherung*) instructs the

the DI requests differ substantially with respect to 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.[22] Consequently, this law introduces further exogenous variation into the process. Further, the decision of the DI offices regarding eligibility usually takes a long time. Thus, the exact moment of receiving money is unanticipated, even in terms of years.[23] Due to these institutional factors, individuals can hardly anticipate whether and when their DI request is approved. Taking together these arguments, the treatment is likely to be randomly assigned once controlled for covariates. Therefore, as outlined in section 1.3.3, the comparison of counterfactual distributions has a causal interpretation. In what follows, I will apply the derived methods to the bivariate distribution of labor income of the two main earners in the household.

1.5.1 Data

I will use annual survey data from the Swiss Household Panel (SHP) from 2004 to 2018. The data contains a sample of roughly 1,500 households with two or more individuals per year and incorporates individual-level information on socio-demographics, health indicators, working hours, income as well as social security benefits. Individuals can be matched to households and vice versa. For the purpose of this analysis, I only keep individuals between 18 and 65 years living in households of two or more persons.[24] Further, I only keep the two biggest contributors to the household's aggregate income.[25] Ranking the individuals by their aggregate income provides

officials opening a request to randomly assign the case to one of 29 advisory institutions. These institutions have to meet 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.

[22]The weighted average of the approval rate is 44% (BSV, 2014). These are the latest statistics for 11 of the 29 advisory institutions. Note that several institutions do not publish those statistics, despite they are obliged to.

[23]Individuals are eligible to receive DIB if they have been unable to work more than 60% of a 42-hours week over the course of one year. The eligibility starts six months after the completion of this year. However, several reasons may extend this waiting time. For instance, individuals working 30 consecutive days lose their eligibility. Note that individuals have to submit a claim to get payments, there is no automatic process to kick off the process.

[24]Note that single households may respond differently, however, the focus lies on the common response within households.

[25]Aggregate income is defined as the sum of labor income, private pension income, retirement pension, unemployment benefits, disability benefits and welfare benefits.

the opportunity to separately analyze whether the response is different depending on who bears the shock. For the remainder of this section, I will refer to these two individuals as *main earner* and *second earner*. On average, the main earner gains 104'143 CHF per year. The second earners contribute considerably less (41'063 CHF). Overall, a fraction of 1.9% of all main earners receive DIB. If so, individuals receive on average 17'247 CHF of DI payments. Tables 1.8 and 1.9 as well as Figure 1.3 in Appendix 1.C provide a detailed description of the variables and a plot of the marginal distributions of labor income, respectively.

1.5.2 Model

As labor supply may depend on a long list of factors. Thus, it is crucial to include demographic variables in the distribution regression model. Similarly, individual labor supply may be constrained by poor labor market positions for older or low skilled individuals. Probably, households with children or located far from city centers have fewer options to increase their labor supply. Hence, X includes the following regressors: gender, age, education, self-rated health status, number of days in the hospital over the last 12 months, an indicator for chronic diseases, the number of children living in the household, an indicator for living in a center or suburban area and a treatment indicator for receiving DI benefits.[26] Apart from the number of children and the residence indicator, the variables vary across spouses.

The household's response likely depends on the amount of DIB. To account for this, I define three different treatments by two equal splits in the distribution of DIB. Specifically, the treatments are newly receiving (i) less than 12'000 CHF, (ii) between 12'000 CHF and 24'000 CHF and (iii) above 24'000 CHF for the main earner.[27] As a results, I will include three different treatment dummy variables, T_1, T_2, T_3, in the regression model. Importantly, these dummies only switches to one (and remains one) if an individual newly starts to receive DIB. Subsequently, I will refer to these treatments as receiving low, average or high DIB. Using these treatment dummies, one can define four counterfactual distributions to evaluate how DI benefits affect the bivariate distribution of labor income. In particular, the covariate matrix is modified to have either all entries of a treatment

[26]Education is measured in years and income is measured in CHF. Unfortunately, the data contains a lot of missings with respect to the health variables. Leaving out these covariates almost doubles the number of observations up to 43'637. Table 1.12 shows that the results do not crucially depend on these variables.

[27]The corresponding values for the second earner are 13'100 CHF and 23'800 CHF.

dummy equal to zero or one. By only varying the covariate distribution and leaving the conditional distribution constant, the resulting counterfactuals can be written as

$$F_{Y\langle\cdot|T\rangle}(t) = \int_{\mathcal{X}_T} F_{Y|X}(t|X) dF_{X_T}(X)$$

where T takes on a value between 0 or 3 referring to no, low, average or high DIB. Note that the support condition is trivially fulfilled as only a binary variable is modified. The comparison of the counterfactuals and their implied characteristics may be viewed as a ceteris paribus change of the treatment. In what follows, I will focus on the treatment of the main earner, arguably the most substantial shock to the household.[28] In particular, it is constructive to look at the conditional quantile function of the second earner, provided that the main earner is at a specific quantile of labor earnings. Formally, the CDF of the second earner at the median of the main earners income $Q_{50}(Y_1)$ is $F_{Y_2\langle\cdot|Y_1=Q_{50}(Y_1),X\rangle}(t) = \frac{F_{Y\langle\cdot|X\rangle}(Q_{50}(Y_1),Y_2,t)}{F_{Y\langle\cdot|X\rangle}(Q_{50}(Y_1),\infty,t)}$. Taking the left inverse of this CDF yields the conditional quantile function. I use the logistic link function and a grid of 25×25 thresholds as a benchmark.

1.5.3 Results

Table 1.2 presents a summary of the average treatment effects. These averages are directly implied by the CDFs presented in Figure 1.2 in Appendix 1.C. The main earner entering DIB reduces both, the main earner's and the second earner's income. The magnitude of the effects is sizeable: A reduction in labor income up to 68'341 CHF for the main earner and up to 15'591 CHF for the second earner. Considering the main earner, this result confirms existing findings for Spain and the US (Autor et al., 2016; Gelber et al., 2017; Marie and Castello, 2012). Note that the estimates in Table 1.2 do not directly translate to elasticities of labor supply as the treatment does not reflect a marginal change. Further, the average effects show that second earners do not compensate for the loss in household income as suggested in Autor et al. (2019). Possibly, instead of raising labor supply, second earners spend an increased amount of time taking care of their spouses (Lee, 2020). Another explanation could be that the average household is well insured by the social security system and 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

[28]The corresponding results regarding the treatment of the second earner as well as various robustness checks can be found in Appendix 1.C.

Table 1.2: Disability Main Earner, Average Treatment Effects

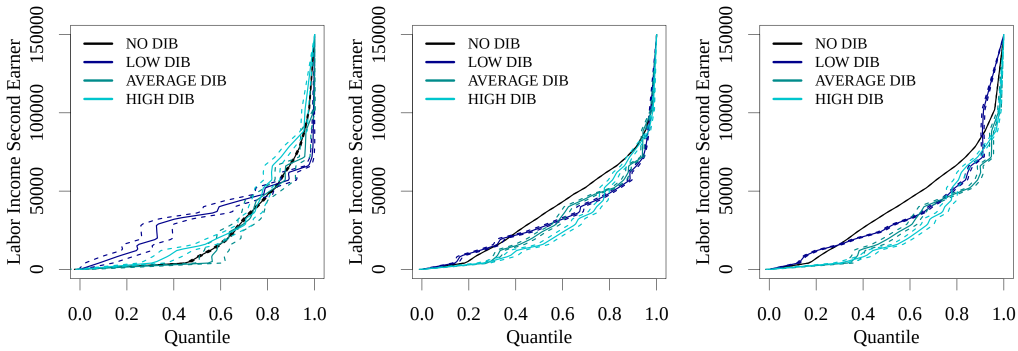
	No DIB (CHF)	Low DIB (Δ CHF)	Avg DIB (Δ CHF)	High DIB (Δ CHF)
ME: Mean	98748	-11969⋆	-68341⋆	-63276⋆
	(831)	(16069)	(3897)	(10116)
ME: SD	55214	26607⋆	-22037⋆	-19275⋆
	(1590)	(25424)	(4920)	(6878)
SE: Mean	39156	-66	-15020⋆	-15591⋆
	(633)	(7408)	(3703)	(5397)
SE: SD	33322	6079	-7433⋆	-5433⋆
	(458)	(9286)	(2694)	(2810)
Corr	0.361	0.26	-0.737⋆	-0.486⋆
	(0.017)	(0.403)	(0.239)	(0.514)

Notes: N = 21'943. Standard deviations are reported in brackets. The stars represent significance on the 5%-level. The standard errors have been estimated using a clustered bootstrap technique with replacement and 500 draws. Apart from the correlation, the values represent amounts in CHF.

that in Switzerland DI benefits depend on previous income only up to a degree (Leisibach et al., 2018, figure on p. 49). Figure 1.4 in Appendix 1.C presents the treatment effects across the distribution of labor income. Remarkably, the absolute treatment effects are close to being constant across the univariate income distribution of main and second earners. Thus, in relative terms, low-income individuals face a substantially higher reduction of labor income. Finally, we take the correlation between the main and second earners income as a simplistic measure of the dependence structure. Table 1.2 indicates that the latter changes substantially across the counterfactual distributions. Mainly, this change is explained by the large drop of the main earner's income for recipients of above-average DIB.

Digging deeper into the behaviour of second earners, Figure 1.1 presents the quantile function of the second earners, conditional on the main earner's quantile of labor income. For most second earners, the quantile function is shifted to the right, which reflects a negative impact on labor income. However, the left panel of Figure 1.1 suggests that the partners of low-income main earners extend their labor supply in response to the treatment. In particular, this is true for households who receive the lowest amounts of DIB because the dark blue quantile function (low DIB) is well above the black line (no DIB). Likely, the low payments do not cover the financial loss incurred by the shock. Thus, second earners increase their labor supply to

Figure 1.1: Quantile Function Second Earner



(a) Quantile ME: 0.1 (b) Quantile ME: 0.5 (c) Quantile ME: 0.9

Notes: N = 21'943. All panels present conditional quantile functions of the second earner depending on how the main earner is treated. The dotted lines represent uniform 95%-Confidence bands computed according to Algorithm 1 in Chernozhukov et al. (2019).

secure the financial stability of the household. Spouses of average or high-income individuals respond differently. Presumably, these households own sufficient assets to bear the shock without having to work more hours. The middle and right panel of Figure 1.1 indicate that this relationship holds irrespective of benefit size. As suggested by Lee (2020), it could be that spouses take care of their disabled partners and the household and thus reduce their labor supply. Taking together, only spouses of low-income earners that receive little DIB seem to respond as suggested by Autor et al. (2019). Considering all the evidence, the present application confirms previous findings and contributes to the literature with two insights: Spouses labor supply response crucially depends on (i) the main earner's income and (ii) on the amount of received DIB. Further, the results relate to the on-going debate on how the poor are exposed to economic shocks and the marginal propensity to consume (e.g. Blundell et al., 2016; Kaplan et al., 2020; Misra and Surico, 2014).

1.6 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, there is no necessity to parametrically specify marginal distributions or dependence structures.

This constitutes a major advantage to copula models, which have been widely used to estimate multivariate distributions. In the case of MDR, I show that the assumption on the link function has a negligible effect on the results. MDR thus substantially lowers the risk of misspecification and the ensuing errors. Second, in terms of accuracy, MDR performs slightly better than the copula models. Third, MDR provides a natural framework to study how regressors affect the distribution of outcomes. Making use of counterfactual theory, treatment effects may be defined in a flexible manner. As suggested in section 1.2, this is potentially beneficial to study the dynamics of equity prices or the allocation of labor within households. Finally, the application in section 1.5 illustrates that MDR is able to deepen established findings by analyzing specific quantiles of the outcome distribution.

The framework of MDR offers various avenues for future research. For instance, the theoretical results in section 1.3 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. There are several additional challenges which could be incorporated into MDR models. For example, 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. Finally, fixed effects similar to the model in Chernozhukov et al. (2018) could be included.

1.A Theoretical Results

1.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

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

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

$$\Psi(\beta, t) = E\left[\left(\Lambda\left[P(X_i)'\beta_t\right] - y_i(t)\right)\left(\frac{\lambda\left[P(X_i)'\beta_t\right]P(X_i)}{\Lambda\left[P(X_i)'\beta_t\right]\left(1 - \Lambda\left[P(X_i)'\beta_t\right]\right)}\right)\right]. \tag{1.A.2}$$

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 $\Lambda\left[P(X_i)'\beta_t\right] \to F_{Y|X}(t)$ 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 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 (1.A.2) 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 $\Lambda\left[P(X_i)'\hat{\beta}_n(t)\right] \xrightarrow{a.s.} y_i(t)$ 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 $P\left(\left|\left|\frac{1}{n}\sum_{i=1}^n \frac{\xi_i}{\bar{\xi}}\psi_{\hat{\beta}^\circ_n,t}\right|\right|_\infty > \eta|\mathcal{X}_n\right) \xrightarrow{a.s.} 0$ which is equivalent to the desired statement in the second part of the condition.[29] 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). Table 1.3 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 $\frac{\lambda[P(X_i)'\beta_t]P(X_i)}{\Lambda[P(X_i)'\beta_t](1-\Lambda[P(X_i)'\beta_t])}$ being uniformly bounded on $b \in R^{dx}$, all these link functions satisfy this condition (see Chernozhukov et al., 2013, p. 2263). ■

[29]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).

1.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 1.3. These are defined as $\frac{\partial}{\partial(b',t)}\Psi(b,t) = [J(b,t), R(b,t)]$ where $\Psi(\beta,t) = E\Big[(\Lambda\left[P(X_i)'\beta_t\right] - y_i(t)) \left(\frac{\lambda[P(X_i)'\beta_t]P(X_i)}{\Lambda[P(X_i)'\beta_t](1-\Lambda[P(X_i)'\beta_t])}\right)\Big]$.

Table 1.3: Parametric Link Functions

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

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

1.B Simulation

1.B.1 Misspecification

As outlined above, copula models are at risk to be misspecified. Considering a Gumbel, Frank and Joe copula specification, the same simulation study as in section 1.4 is performed. The data size is set to 1'000 observations to get comparable results. For the nonparametric copula estimator, the $CMSE$ using a Gaussian kernel is compared to same criterion using a Triangular or Epanechnikov kernel. Apart from the logistic link function, the MDR estimator is computed using a probit or C-log-log link function. Note that the choice of the kernel for the nonparametric copula as well as the choice of the link function for the MDR procedure do not correspond to a classic misspecification problem. Rather, they serve as robustness checks. Overall, the results for unconditional copula model and the MDR estimator are only marginally affected by the respective choices. However, the MDR estimator clearly provides the best fit. Further, the results in Table 1.5 show that the assumption on the copula itself is important. Apart from the Frank copula which seems to perform slightly better than the correctly specified model, misspecification may scale the prediction error up to a factor of 12. The nonparametric estimator performs worse with Triangular or Epanechnikov kernels. There, the error may be about 6 times larger than with the Gaussian kernel.

1.B.2 Scaling P(X)

Considering a richer set of covariates may improve the performance of the MDR estimator. The data generating process outlined in section 1.4 is too simplistic to serve as example in this case. Thus, I introduce Y_{S2} which is generated as follows by taking Y_S and changing the averages. Formally, this is

$$Y_{S2} \sim N\left(\begin{pmatrix} \frac{1}{x} \\ \exp(x) \end{pmatrix}, \begin{pmatrix} 1 & 2\cdot\Phi(2x)-1 \\ 2\cdot\Phi(2x)-1 & 1 \end{pmatrix}\right).$$

It suffices to introduce further non-linearity in X. Any DGP fulfilling this condition would produce similar results. A grid of 25×25 threshold is used. I extend the matrix $P(X)$ by considering indicators for quintiles (model (2) in Table 1.6) and centiles (3) of X, interaction of those with a linear term ((4) and (5), respectively) and by additionally adding polynomials up to the 10^{th} degree ((6) and (7)). Table 1.6 provides the estimation results for Y_S2 considering all different $P(X)$. Overall, the performance may be improved by a factor up to 56. Note that in cases with a small number of observations, large $P(X)$ may overfit and thus provide less accurate results.

1.B.3 Bandwidth Simulation

Using a grid of 15×15 values in .01 to .99, the DGP's defined in section 1.4 are estimated with varying bandwidths. Note that I use two different sets of bandwidths for the reduced and complex DGP, respectively. Mainly, the reason is that the $CMSE$ in the case of Y_2 is seemingly unaffected by large changes in the bandwidth. This can be seen in Table 1.7. Depending on the datasize, a bandwidth between 0.005 and 0.2 yields the best results. Thus, I chose to use a bandwidth of 0.125 for the models estimated in section 1.4.

Table 1.4: IMSE for Copula and DR models, No Variance Correction

Obs.	Statistic	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
100	IMSE	0.0139	0.0120	0.0049	0.0035	0.0049
100	IVAR	0.0000	0.0000	0.0003	0.0001	0.0004
100	IBIAS	0.0139	0.0120	0.0046	0.0034	0.0045
1000	IMSE	0.0074	0.0060	0.0025	0.0023	0.0033
1000	IVAR	0.0000	0.0001	0.0021	0.0021	0.0022
1000	IBIAS	0.0074	0.0059	0.0003	0.0002	0.0011
10000	IMSE	0.0035	0.0026	0.0035	0.0038	0.0032
10000	IVAR	0.0001	0.0002	0.0035	0.0038	0.0029
10000	IBIAS	0.0035	0.0023	0.0001	0.0000	0.0003

Notes: The values represent averages over 10'000 replications. (1): Unconditional Normal Copula Model, (2): MDR including only a constant, (3): Nonparametric Copula Model, (4): Parametric Normal Copula Model, (5): MDR. The statistics are averages over the 25×25 cells.

Table 1.5: IMSE for Copula and DR models under Misspecification

Model	Specification	IMSE	IVAR	IBIAS
Unconditional Copula Model	Correctly Specified	41.4469	0.0008	41.4461
	Gumbel	45.1472	0.0009	45.1464
	Frank	41.7315	0.0008	41.7306
	Joe	44.2828	0.0009	44.2819
Parapemtric Copula Model	Correctly Specified	3.3107	0.0157	3.2950
	Gumbel	40.1012	0.0048	40.0964
	Frank	1.6616	0.0186	1.6430
	Joe	40.0975	0.0045	40.0930
Non-Parametric Copula Model	Gaussian	1.7201	0.0171	1.7030
	Triangular	10.1170	0.0241	10.0929
	Epanechnikov	8.1835	0.0226	8.1609
Distribution Regression	Logit	0.5023	0.0153	0.4870
	Probit	0.2103	0.0158	0.1945
	Cloglog	0.7792	0.0140	0.7651

Notes: The values represent averages over 1′000 replications. All models are estimated using 1′000 observations and a grid of 25 × 25 values.

Table 1.6: Scaling P(X)

Obs.	Statistic	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
100	IMSE	2256.7626	34.6436	129.2788	9386.1790	9374.2562	10980.7354	10973.5831
100	IVAR	0.0830	0.0118	0.0260	0.2215	0.2835	0.3968	0.4748
100	IBIAS	2256.6796	34.6319	129.2528	9385.9575	9373.9727	10980.3386	10973.1083
1000	IMSE	157.4353	4.2129	1.8942	7.4140	4.6762	59.4052	62.9836
1000	IVAR	0.0328	0.0306	0.0517	0.0573	0.0683	0.3034	0.3826
1000	IBIAS	157.4025	4.1823	1.8425	7.3567	4.6079	59.1019	62.6009
10000	IMSE	33.8279	4.3682	5.2847	7.1530	7.0639	0.6019	0.6332
10000	IVAR	0.8384	0.0474	0.1024	0.0938	0.1221	0.4061	0.4413
10000	IBIAS	32.9894	4.3207	5.1823	7.0592	6.9418	0.1958	0.1919

Notes: The values represent averages over 100 replications. $P(X)$ includes the following. (1): Only a constant and a linear term for X, (2): (1) and indicators for each quintile of X, (3): (1) and indicators for each centile of X, (4): (2) and all interactions of the quintiles with a linear term of X, (5): (3) and all interactions of the centiles with a linear term of X, (6): (4) and polynomials up to the 10^{th} degree, (7): (5) and polynomials up to the 10^{th} degree.

Table 1.7: Bandwidth Simulation, $X \sim N(0,1)$

Bandwidth	Non-Parametric Copula			Parametric Copula		
Datasize	100	1000	10000	100	1000	10000
0.05	1070.3448	20.3472	0.0372	20.4180	0.3523	0.0334
0.1	1173.8093	4.3634	0.0363	20.1498	0.1637	0.0328
0.15	836.7580	2.8825	0.0353	16.2091	0.1514	0.0315
0.2	803.0485	0.9678	0.0340	11.3948	0.1690	0.0297
0.25	784.6399	0.0636	0.0326	10.5589	0.2052	0.0277
0.3	739.1424	0.1354	0.0310	10.3940	0.2617	0.0258
0.35	667.7641	0.1515	0.0294	10.4526	0.3354	0.0240
0.4	578.9029	0.1539	0.0278	10.7782	0.4259	0.0225
0.45	77.2748	0.1614	0.0263	10.6378	0.5024	0.0214
0.5	49.4032	0.2882	0.0250	10.3061	0.6284	0.0208

Notes: The values represent averages over 1000 replications.

1.C Application Results

Table 1.8: Descriptive Statistics, Individuals living in Couples

	Mean	Median	SD	Min	Max	N
Labor Income (CHF)	72603	66587	69004	0	3000000	43,886
Age	45	46	11	18	65	43,886
Number of Years of Education	14	12	3	9	21	43,886
Number of Children in the HH	.88	0	1.1	0	8	43,886
Female	.49	0	.5	0	1	43,886
Chronic Disease	.33	0	.47	0	1	43,886
Self-rated Health (1: well, 5: bad)	1.9	2	.62	1	5	43,886
Hospital Days over the last year	.85	0	5.3	0	330	43,886
Living in a Center	.53	1	.5	0	1	43,886
Received DI Benefits	.019	0	.14	0	1	43,886
Started Receiving DI Benefits	.011	0	.11	0	1	43,886
DI Benefits (CHF) if > 0	17587	18000	8727	90	32600	840

Notes: This table only includes observations, for which all variables of both spouses are non-missing.

Table 1.9: Differences within Couples

	Main Earner		Second Earner	
	Mean	SD	Mean	SD
Labor Income (CHF)	104143	(79067)	41063	(35806)
Age of Individual	46.9	(10.3)	42.9	(11.8)
Number of Years of Education	14.5	(2.95)	13.5	(2.95)
Female	.168	(.374)	.812	(.39)
Chronic Disease	.319	(.466)	.336	(.472)
Self-rated Health (1: well, 5: bad)	1.9	(.604)	1.94	(.642)
Hospital Days over the last year	.721	(4.68)	.969	(5.91)
Received DI Benefits	.0159	(.125)	.0224	(.148)
Started Receiving DI Benefits	.0113	(.106)	.0114	(.106)
DI Benefits (CHF) if > 0	17247	(9251)	17829	(8336)

Notes: The number of observations is 21'943. This table only includes observations, for which all variables of both spouses are non-missing.

Table 1.10: Disability Main Earner, All Statistics

	No DIB (CHF)	Low DIB (CHF)	Avg. DIB (CHF)	High DIB (CHF)
Mean ME	98748	-11969⋆	-68341⋆	-63276⋆
	(831)	(16069)	(3897)	(10116)
Median ME	93731	-14692⋆	-51731⋆	-51731⋆
	(686)	(10664)	(8986)	(14475)
SD ME	55214	26607⋆	-22037⋆	-19275⋆
	(1590)	(25424)	(4920)	(6878)
Mean SE	39156	-66	-15020⋆	-15591⋆
	(633)	(7408)	(3703)	(5397)
Median SE	36400	-7812⋆	-16900⋆	-20848⋆
	(879)	(4832)	(7722)	(6912)
SD SE	33322	6079	-7433⋆	-5433⋆
	(458)	(9286)	(2694)	(2810)
Cov	6.6e+08	1.3e+09⋆	-9.9e+08⋆	-7.9e+08⋆
	(5.0e+07)	(1.5e+09)	(1.8e+08)	(3.2e+08)
Corr	0.361	0.26	-0.737⋆	-0.486⋆
	(0.017)	(0.403)	(0.239)	(0.514)

Notes: N = 21'943. The stars represent significance on the 5%-level. The standard errors are reported in brackets and have been estimated using a bootstrap technique with replacement and 500 draws.

Table 1.11: Disability Second Earner, All Statistics

	No DIB (CHF)	Low DIB (CHF)	Avg. DIB (CHF)	High DIB (CHF)
Mean ME	98748	86780⋆	30407⋆	35473⋆
	(831)	(16081)	(3891)	(10017)
Median ME	93731	79039⋆	42000⋆	42000⋆
	(686)	(10731)	(9083)	(14535)
SD ME	55214	81821⋆	33178⋆	35940⋆
	(1590)	(25480)	(4477)	(6665)
Mean SE	39156	39090⋆	24136⋆	23565⋆
	(633)	(7534)	(3811)	(5444)
Median SE	36400	28588⋆	19500⋆	15552⋆
	(879)	(4940)	(7793)	(6866)
SD SE	33322	39401	25888⋆	27888⋆
	(458)	(9311)	(2678)	(2895)
Cov	6.6e+08	2.0e+09⋆	-3.2e+08⋆	-1.3e+08⋆
	(5.0e+07)	(1.5e+09)	(1.7e+08)	(3.1e+08)
Corr	0.361	0.621	-0.376⋆	-0.126⋆
	(0.017)	(0.403)	(0.242)	(0.518)

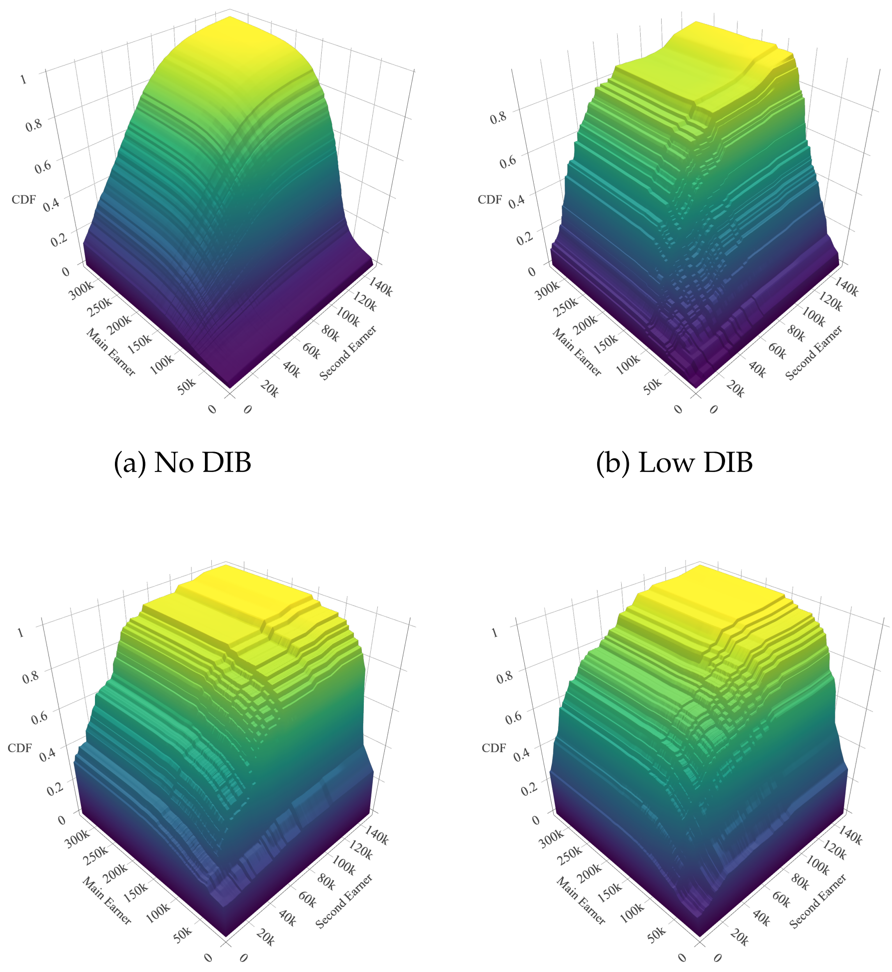
Notes: N = 21’943. The stars represent significance on the 5%-level. The standard errors are reported in brackets and have been estimated using a bootstrap technique with replacement and 500 draws.

Table 1.12: Treatment Effects, All Statistics, No Health Covariates

	Disability Main Earner			Disability Second Earner		
	T=0	T=1	TE	T=0	T=1	TE
Mean ME	97664	50652	-47012⋆	97263	83713	-13550⋆
	(1402)	(7997)		(1417)	(8897)	
Median ME	91000	43900	-47100⋆	90695	73315	-17380⋆
	(887)	(8691)		(832)	(8505)	
Var ME	3.3e+09	2.8e+09	-5.0e+08⋆	3.3e+09	3.3e+09	-3.1e+07⋆
	(3.7e+08)	(1.0e+09)		(3.8e+08)	(1.0e+09)	
Mean SE	36117	24521	-11596⋆	36148	17697	-18451⋆
	(705)	(4155)		(710)	(5709)	
Median SE	32110	15689	-16421⋆	33000	6300	-26700⋆
	(1207)	(5455)		(1268)	(5234)	
Var SE	9.9e+08	7.5e+08	-2.4e+08⋆	9.9e+08	6.4e+08	-3.5e+08⋆
	(4.1e+07)	(1.9e+08)		(4.1e+07)	(3.4e+08)	
Cov	7.4e+08	6.2e+08	-1.2e+08⋆	7.5e+08	4.8e+08	-2.7e+08⋆
	(6.9e+07)	(3.3e+08)		(7.0e+07)	(4.7e+08)	
Corr	0.41	0.42	0.02⋆	0.41	0.33	-0.08⋆
	(0.02)	(0.16)		(0.02)	(0.21)	

Notes: N = 21’943. The stars represent significance on the 5%-level. The standard errors have been estimated using a bootstrap technique with replacement and 100 draws.

Figure 1.2: Bivariate Distribution of Labor Earnings



(a) No DIB

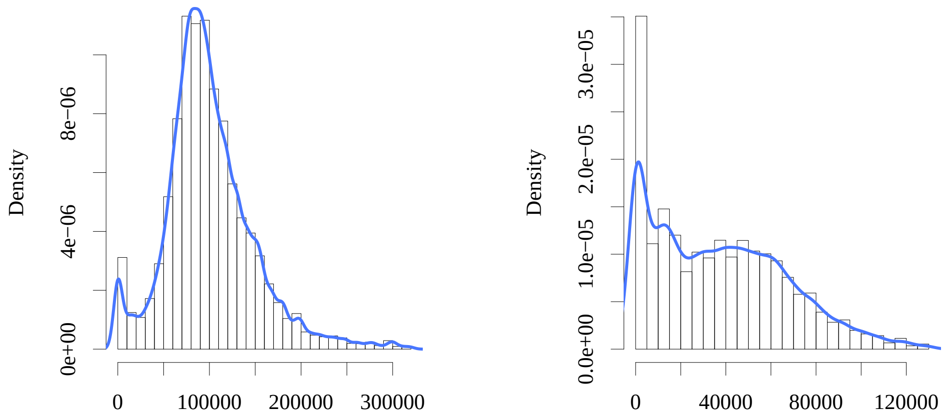
(b) Low DIB

(c) Average DIB

(d) High DIB

Notes: The 3D Figures present the bivariate CDF of spouses labor supply. Each one of the panels refers to a counterfactual distribution when all individuals would receive (a) no DIB, (b) low DIB, (c) average DIB and (d) high DIB.

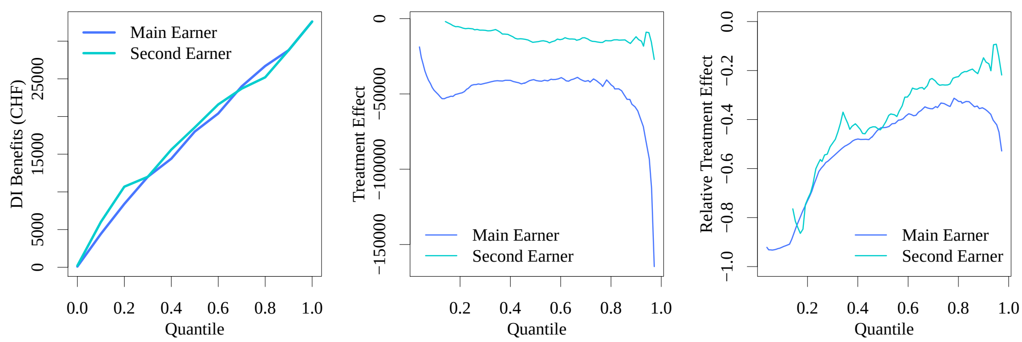
Figure 1.3: Histogram Labor Income



(a) Main Earner (b) Second Earner

Notes: The histogram presents 30 distinct bins and the density has been estimated using a Gaussian kernel. The top 1% of both distributions is not plotted to improve readability.

Figure 1.4: Distribution of DI Benefits and Quantile Effects



(a) Descriptive (b) QTE (c) Relative QTE

Notes: For all panels, the x-axis represents the quantiles of labor earnings. The left panel presents the deciles of DI benefits only for the recipients, N = 423 for main earners and N = 250 for the second earners. In extension to the results in Table 1.2, the middle panel provides the difference of the quantile functions implied by the counterfactual distributions. The right panel descriptively presents these quantile effects divided by the corresponding value of the quantile.

Chapter 2

Censored Distribution Regression

joint with **Blaise Melly**

Abstract

When researchers aim to estimate the effect of explanatory variables on duration, the most popular approach is the Cox proportional hazard estimator. This estimator is very flexible concerning the baseline hazard but restricts the covariates' effect to be constant over time. Distribution regression is a generalization of the Cox model that naturally allows the coefficients to change with time. Typically, survival times are censored, making the inclusion of time-varying effects challenging. This paper introduces a novel censored distribution regression estimator based on maximum likelihood estimation (CDR-MLE) that overcomes this difficulty. We assume that the censoring time is independent of the event time conditional on the covariates, which is the standard type of censoring in the literature. The CDR-MLE estimator is numerically identical to the Kaplan-Meier estimator when the covariates include only a constant and converges to the Cox coefficients when the coefficients are time-constant. It consistently estimates the conditional survival distribution for discrete, continuous, and mixed duration. We show weak convergence of the censored distribution regression coefficient process and the conditional distribution process to centered Gaussian processes. In addition, we complement the literature by providing undiscovered links to an existing estimator. Simulation studies show the good finite-sample properties of the CDR-MLE estimator. Finally, we apply our estimator to unemployment duration data and find that a

Acknowledgements: We would like to acknowledge the financial support from the Swiss National Science Foundation for the grant 165621.

longer benefit duration has little effect on short unemployment spells but a strong and significant effect on long-term unemployment.

2.1 Introduction

A policy intervention that reduces long-term unemployment (i.e., the upper tail of the distribution) is often more appreciated than an intervention that shifts the distribution's lower tail, even if both interventions' average effects are identical. In this example, like in many others, it is essential to assess the distributional effects of policy variables on durations. The most popular approach in the field is the Cox (1972) proportional hazard estimator, which assumes that the covariates' effects are homogeneous. Consequently, researchers can only address average effects on durations. In this paper, we promote two novel estimators that generalize the Cox model. In particular, the estimated regression coefficients converge to the same value when the coefficients are time-constant, but they capture the underlying heterogeneity when the effects vary over time.

Censoring is an ubiquitous phenomenon in duration data. In the absence of covariates, the standard estimator of the survival function was suggested by Kaplan and Meier (1958). This estimator imposes independence between the censoring times and the event times. Under this assumption, it can be interpreted as the maximum likelihood estimator in the unrestricted model. Throughout this paper, we will assume standard censoring in the presence of covariates, i.e., that the censoring times are independent of the event times conditional on the covariates.

As noted by McCullagh (1980) and Doksum and Gasko (1990), the Cox proportional hazard model implies that a binary regression with a complementary log-log link function can estimate the conditional distribution function. We extend this idea by estimating not one but many binary regressions. In particular, we suggest estimating a binary regression at each threshold of the outcome. This estimation strategy is known as distribution regression and presents a natural generalization of the homogeneous proportional hazard model. Accordingly, it allows testing the validity of the Cox model. In the absence of censoring, the distribution regression estimator is well behaved without further adjustments. In the presence of censoring, we will introduce weighted pseudo-observations in the spirit of Efron (1967). The resulting estimator is numerically identical to the unconditional Kaplan-Meier estimator if the only regressor is a constant and equals the Kaplan-Meier estimator within cells if the model is fully saturated. Thus, our estimator is also a strict generalization of the Kaplan-

Meier estimator in the presence of covariates. In both cases, unconditional or conditional on covariates, our estimator can be applied without modification to discrete, continuous, and mixed discrete-continuous outcomes. As each of the binary regressions is estimated by maximum likelihood, we refer to this novel estimator as CDR-MLE, where CDR stands for censored distribution regression.

Most closely related, Peng and Huang (2007) consider the same model and suggest an estimator assuming a martingale structure of the data. In contrast to the CDR-MLE estimator, their strategy falls short of handling discrete or mixed outcomes. As duration data frequently incorporates such outcome variables, we extend their existing estimator to accommodate all types of durations. We will refer to the resulting estimator as the CDR-Martingale estimator. Note that the asymptotic results follow analogously to the initial estimator, which is why we abstract from the proofs in this paper. With our modification, the novel estimator becomes numerically identical to the Kaplan-Meier estimator in saturated models. In this regard, it closely matches the properties of the CDR-MLE estimator. To assess their qualitative equivalence, we compare the performance of both estimators in various simulation studies. Thus, in addition to introducing the CDR-MLE estimator, we complement the literature by linking concepts and consolidating them into a unified framework.

The estimation strategy of the CDR-MLE estimator adopts the concepts of distribution regression as follows. We start estimating the conditional cumulative distribution function (CDF) recursively from the lower tail of the distribution. While no observation has been censored before the survival time we consider, the conditional CDF can be estimated consistently via standard binary regressions. After we have crossed the first censored observation, we redistribute the weight of the censored observations to two pseudo-observations, one below the threshold and one above it in the spirit of Efron (1967). Then we can run a standard weighted binary regression. The weights are functions of the previously estimated binary regressions, requiring estimating the regressions recursively starting from the bottom. For discrete outcomes, the weights are also functions of the binary regression results at the threshold. Therefore, we suggest iterating the procedure at each threshold until convergence.

Our CDR-MLE estimator is a Z-estimator with an estimating equation that generalizes the 'self-consistency' expression for the Kaplan-Meier estimator. We use empirical process arguments to show consistency and weak convergence of the CDR coefficient process. By the functional delta method, it follows that the conditional distribution function process is also consistent and converges to a mean-zero Gaussian process. Using these results, we can test functional hypotheses and compute uniform confidence bands using resampling procedures. In particular, using a complementary log-log link function, we can address the validity of the Cox proportional hazard model by testing if the coefficients are constant across time.

The cost of the CDR model's flexibility is the high-dimensionality of the coefficients and the difficulty to interpret their value, except for the homogeneous reference case of the Cox model. However, these coefficients ultimately determine the conditional and unconditional distribution function of the outcome. Thus, we suggest reporting more intuitive parameters such as the quantile functions of conditional, unconditional, or counterfactual distributions. In the application, we illustrate the advantage of such functionals. In particular, we report the quantile treatment effect function, which is the difference between the quantile function that we would observe if everyone had a long period of benefits and the quantile function that we would observe if everyone had a short period of benefits.

We perform simulation studies to analyze the finite-sample performance of both CDR estimators. Already in samples of moderate size, the asymptotic results describe the behavior of the estimators well. Further, our simulations suggest that the CDR-MLE and CDR-Martingale estimator provide qualitatively equivalent results. When the effects are homogeneous, as expected, the CDR estimator is less precise than the Cox estimator, but the difference remains moderate. Taking the average variance of the CDR estimator from the 10th to the 90th quantile of the outcome, we find an increase of 11% compared to the Cox model. When the effects are heterogeneous, the Cox model's bias quickly dominates the CDR estimator's slightly higher variance. In particular, the higher bias causes the mean square error to increase by a factor of five already in moderate sample sizes.

An application illustrates the usefulness CDR. We examine how an increase in potential benefit duration (PBD) affects unemployment spells. Theoretical search models predict that job search behavior varies throughout unemployment, implying heterogeneous effects of longer PBD. In particular, faced with the upcoming exhaustion of benefits, individuals intensify their efforts to find a job. Our results suggest that, on average, a longer PBD of 9 months decreases the likelihood of finding a job by 5%, which is equivalent to an increase of roughly 2.5 weeks in duration. The effect is much more substantial for the long-term unemployed. For these individuals, a longer PBD increases duration by up to 14 weeks. Our findings are in line with the suggested time-varying behavior of job seekers.

The rest of the paper is organized as follows. In Section 2.2 we discuss previous work with a particular focus on methods suggested to analyze heterogeneous effects of covariates on duration data. In Section 2.3 we define the model that we study, state the assumptions, and show that the parameters are identified. In Section 2.4 we derive the theoretical properties of the estimator. Finally, in Section 2.5 we provide the results of the Monte Carlo simulations, and in Section 2.6 the results of the application to real data.

2.2 Related work

The CDR estimators at hand simplify (i) to the Kaplan-Meier estimator in the absence of covariates, (ii) to distribution regression in the absence of censoring, and (iii) to the Cox estimator in the absence of heterogeneity. Thus, we discuss these three strands of literature. In addition, we also discuss the censored quantile regression and the additive hazard models, which are two alternative methods to estimate heterogeneous effects of covariates on the conditional distribution of a duration. We refer the reader to Peng and Huang (2007) for an extensive discussion of the CDR-Martingale estimator and focus on the novel CDR-MLE estimator in the following.

Given the good properties of the Kaplan-Meier estimator, Beran (1981) suggested a nonparametric conditional version of this estimator using either a nearest neighbor or a kernel approach. Dabrowska (1987) and Dabrowska (1989) established uniform consistency and weak convergence of these estimators. The obvious drawback of this local approach is the low rate of convergence and the curse of dimensionality when the covariates' dimension is large. We avoid this drawback and obtain $\sqrt{n}$-consistent estimators at the price of imposing a semiparametric model similar to the Cox model.

Distribution regression, as defined in Chernozhukov et al. (2013), consists of the application of a sequence of binary regressions to estimate the whole conditional distribution of an outcome given covariates. This approach has already been suggested by Williams and Grizzle (1972) for ordered outcomes. Foresi and Peracchi (1995) apply this procedure to estimate the conditional distribution of a continuous outcome at a finite number of thresholds. Chernozhukov et al. (2019) show that distribution regression nests all generalized linear regression model and apply it to count data. One advantage of distribution regression is that it handles continuous, discrete, or mixed dependent variables without any special adjustment.

Survival data are typically censored. If the censoring point is the same for all observations, distribution regression is still consistent below the censoring point without modification.[1] However, the censoring time in survival analysis is typically random and observed only for censored observations. Jung (1996) is the first to consider the estimation of distribution regression with random censoring. He assumes that the censoring times are independent of the covariates. This allows him to estimate the distribution of the censoring time with the unconditional Kaplan-Meier estimator. In a second step, the coefficients can be estimated with binomial regressions weighted by the inverse probability of censoring. Scheike and Zhang (2007) and Scheike et al. (2008) suggest a very similar approach. Sant'Anna (2016) and Delgado et al. (2019) suggest estimators based on integrals of the multivariate Kaplan-Meier estimator of Stute (1993). All these approaches assume that the censoring times and the event times are unconditionally independent. The covariates are not allowed to affect the distribution of the censoring variable. As a consequence, in a fully saturated regression model, these estimators are not equal to the Kaplan-Meier estimator within the cells. In contrast, we assume standard censoring, i.e., the censoring times and the event times are independent conditionally on the covariates. Our estimator is numerically equal to Kaplan-Meier within the cells for fully saturated models.

Instead of modeling the conditional distribution of the duration given covariates, we could model the conditional quantile function. This approach has first been suggested by Koenker and Geling (2001) and Koenker and Bilias (2001) and has been extended to random censoring by Portnoy (2003) and Peng and Huang (2008). In a nonparametric framework, it does not matter if we model the quantile or distribution function. In parametric settings, the difference between these models consists of the way heterogeneity is allowed. With quantile regression, two observations at the same conditional quantile are assumed to have the same coefficients. With distribution regression, two observations at the same duration are assumed to have the same coefficients. In general, it is not clear what approach is more desirable. We can note, however, that quantile regression does not allow for a time-varying hazard rate. It means that quantile regression does not generalize the Cox model while distribution regression does it. Also, the linearity assumption for the conditional quantiles underlying quantile regression is highly implausible for discrete data. For instance, the linear proportional hazard model does not have linear conditional quantiles.

[1] This is the case in the application of Chernozhukov et al. (2013), where wages are censored from below by the minimum wage. The same logic applies if earnings are top-coded.

2.3 Model, Identification and Estimator

Let Y^* denote the scalar outcome variable of interest, typically a duration, and X the $k \times 1$ vector of covariates that includes a constant. In our application in Section 2.6 we are in particular interested in the effect of the maximum duration of unemployment benefits on the length of unemployment spells. The conditional distribution of Y^* given $X = x$ evaluated at a threshold y can be expressed as $F_{Y^*|X}(y|x) = E[1(Y^* \leq y)|X = x]$, where $1(\cdot)$ is the indicator function. Accordingly, we can estimate the conditional distribution at one point with a traditional binary regression. We assume that the link function is known:

Assumption 5 (Link function)**.**

$$F_{Y^*|X}(y|x) = \Lambda\left(x'\beta(y)\right) \qquad \textit{for all } y \in (0, y_U] \textit{ and } x \in \mathcal{X},$$

where $\Lambda(\cdot)$ is a known link function, $\beta(y)$ a $k \times 1$ vector of unknown coefficients, y_U a deterministic constant, and $\mathcal{X}$ the support of X.

For duration data, the complementary log-log link function, $\Lambda(x'\beta(y)) = 1 - exp(-exp(x'\beta(y))$, is of particular interest. The Cox proportional hazard model implies assumption 5 with this link function and the restriction that the slope coefficients (all the elements of $\beta(y)$ except the intercept) are not a function of y. Thus, the DR model generalizes the Cox model by allowing the effect of the covariates to change over time. Other link functions are also of interest too. Bennett (1983) shows that the proportional odds model implies an homogenous DR model with a logistic link function. Peng and Huang (2007) show that the additive hazard model of Aalen (1980) is equivalent to the DR model with the exponential link function $\Lambda(x'\beta(y)) = exp(-x'\beta(y))$.

The DR model is flexible in the sense that, for any given link function, we can approximate the conditional CDF arbitrarily well by using a rich enough set of transformations of the original covariates. In the extreme case when the covariates are all discrete and X is fully saturated, the estimated conditional distribution is numerically equal to the empirical DF in each cell of X for any monotonic link function.

Knowledge of the function $y \mapsto \beta(y)$ implies knowledge of the entire conditional distribution of Y given X. Thus, we can calculate the covariates' effect on the conditional mean or any conditional quantiles of Y. In addition, we can obtain the unconditional distribution of Y by integrating over the covariates as suggested in Chernozhukov et al. (2013). In particular, this procedure allows estimating counterfactual distributions that we would observe if we would change the covariates' distribution. For instance, in our application, we estimate the distributions that we would observe if all individuals received unemployment benefits either during a long or short period. The difference between these two counterfactual distributions summarizes the effect of a policy change.

Let C denote the censoring time. Because of right censoring we do not observe Y^* but $Y = min(Y^*, C)$. $\delta = 1(Y^* \leq C) = 1(Y = Y^*)$ is an indicator variable for being not censored.

Assumption 6 (random sample)**.** *We observe $\{x_i, y_i, \delta_i\}_{i=1}^n$, a random sample of size n from (X, Y, δ).*

Without further assumption, we cannot identify the conditional distribution of Y^*. For instance, if censoring is informative (i.e. a function of Y^* even after conditioning on X), the best we can hope to achieve are bounds for the distribution. In this paper, we assume that censoring is conditionally uninformative:

Assumption 7 (standard censoring)**.**

$$Y^* \perp\!\!\!\perp C|X$$

This assumption is sometimes called standard random right censoring or conditionally independent censoring. It is less restrictive than assuming unconditional independence because it allows the distribution of C to be a function of X.

Before considering this assumption's implications, it is interesting to mention that simpler estimators exist if we make stronger assumptions. For instance, if the censoring time is fixed and observed, which corresponds to the assumption made in Powell (1986), then the standard DR estimator can be applied to estimate the conditional distribution up to $y = C$. This stands in contrast to quantile regression, which requires a different (and considerably more complex) estimator.

If C is not fixed but is always observed, the conditional distribution of Y^* at y can be estimated by running a standard distribution regression on the subsample for which $C > y$. This estimator is consistent under assumption 7.

If C is random but unconditionally independent of Y^*, the distribution of the censoring time can be consistently estimated by the (unconditional) Kaplan-Meier estimator. In a second step, we can estimate the conditional distribution of Y^* simply by reweighting the dependent variable of the binary regression by the probability of being censored. This is similar to inverse probability weighting for missing data. This estimator has been suggested by Jung (1996), Scheike and Zhang (2007), and Scheike et al. (2008).

Our problem is more complex: we cannot identify the distribution of C in a preliminary step. We suggest two estimators based on two different interpretations of the Kaplan-Meier estimator. Both simplify numerically to the Kaplan-Meier estimator when X contains only a constant. The first estimator generalizes the "redistribution of mass to the right" approach of Efron (1967). The second one exploits the martingale approach introduced in survival analysis by Aalen (1975).

2.3.1 Redistribution of Mass to the right

We want to identify $F_{Y^*|X}(y|x) = E[1(Y^* \leq y)|X = x]$. There are three types of observations. First, observations with $\delta_i = 1$ do not cause any complication because $y_i^* = y_i$ is observed in this case. Second, observations with $\delta_i = 0$ and $y_i \geq y$ appear more complicated at the first sight but, even if we do not observe y_i^*, we know that $y_i^* > y$ for these observations, which is all what we need to estimate the CDF at y.[2] Third, observations with $\delta_i = 0$ and $y_i < y$ are truly problematic because we don't know if y_i^* is below or above y. However, Assumption 7 implies that

$$Pr[y_i^* \leq y|\delta_i = 0, y_i \leq y, X = x_i] = \frac{F_{Y_i^*|X}(y|x_i) - F_{Y^*|X}(y_i|x_i)}{1 - F_{Y^*|X}(y_i|x_i)} \equiv w(y, y_i, x_i) \tag{2.3.1}$$

It follows that we can replace each observation with $\delta_i = 0$ and $y_i < y$ with two weighted pseudo-observation: one below y with a weight $w(y, y_i, x_i)$ and another above y with a weight $1 - w(y, y_i, x_i)$. We then run a binary MLE on the expanded sample.

[2]We use the standard convention that $\delta_i = 1(y_i^* \leq c_i)$, which implies that y_i^* is strictly larger than y if it is censored at y.

More formally, note that

$$\begin{aligned}
F_{Y^*|X}(y|x) &= E[1(Y^* \leq y)|X = x] \\
&= E[1(Y^* \leq y) \cdot \delta + 1(Y^* \leq y) \cdot (1-\delta)|X = x] \\
&= E[1(Y \leq y) \cdot \delta + 1(Y^* \leq y) \cdot 1(Y \leq y) \cdot (1-\delta)|X = x] \\
&= E\Bigg[1(Y \leq y) \cdot \delta + 1(Y \leq y) \cdot \\
&\quad (1-\delta)\frac{F_{Y^*|X}(y|x) - F_{Y^*|X}(Y|x)}{1 - F_{Y^*|X}(Y|x)}|X = x\Bigg] \\
&= E\left[1(Y \leq y) \cdot \left(\delta + (1-\delta)\frac{F_{Y^*|X}(y|x) - F_{Y^*|X}(Y|x)}{1 - F_{Y^*|X}(Y|x)}\right)|X = x\right]
\end{aligned} \tag{2.3.2}$$

where the third equality from the observation rule and the fact that $1(Y^* \leq y) \implies 1(Y \leq y)$, and the fourth from equation (2.3.2). The expression on the right-hand side is a function of $F_{Y^*|X}(Y|x)$ but only for values of Y below y. This suggests a recursive algorithm that starts at the bottom of the distribution. Note that $F_{Y^*|X}(y|X_i)$ appears on the left- and right-hand side of 2.3.2. If we solve the sample analog of the equation recursively when X contains only a constant, we obtain exactly the Kaplan-Meier estimator.

With covariates, we suggest estimating recursively $F_{Y^*|X}(y|x)$ with the MLE of a weighted binary regression model. If the weights $w(y,x)$ were known, we would maximize the following likelihood function:

$$\log(L(y)) = \sum_{i=1}^{n} w_i^1 \log(\Lambda(x_i'b)) + w_i^0 \log(1 - \Lambda(x_i'b)) \tag{2.3.3}$$

with

$$\begin{aligned}
w_i^1 &= \delta_i \cdot 1(y_i \leq y) + (1-\delta_i) \cdot 1(y_i \leq y) \cdot w(y, y_i, x_i) \\
w_i^0 &= \delta_i \cdot 1(y_i > y) + (1-\delta_i) \cdot 1(y_i \leq y) \cdot (1 - w(y, y_i, x_i)) \\
&\quad + (1-\delta_i) \cdot 1(y_i > y)
\end{aligned}$$

Note that the observations with $\delta_i = 1$ or with $y_i > y$ enter normally into the likelihood while the observations with $\delta_i = 0$ and $y_i \leq y$ enters with two weighted pseudo-observations. This estimator can be easily implemented with standard software for binary regression.

The weights $w(y, y_i, x_i)$ are actually not known. With the semiparametric assumption 5, these weights can be expressed as

$$w(y, y_i, x_i) = \frac{\Lambda(x_i'\beta(y)) - \Lambda(x_i'\beta(y_i))}{1 - \Lambda(x_i'\beta(y_i))}$$

Since the weights only appear as multiplied with $1(y_i \leq y)$, we need only to estimate the weights for observations with $y_i \leq y$. We estimate the coefficients sequentially, starting from the bottom of the distribution. Thus, we already have an estimator for $\beta(y_i)$ that we denote by $\hat{\beta}(y_i)$. To estimate $\beta(y)$, when the outcome is continuous, we can use the last previously estimated coefficients, which we denote by $\hat{\beta}(y-)$. By continuity of the DR coefficient process, this provides a consistent estimator. When the outcome is discrete, we start from the weights obtained with $\hat{\beta}(y-)$. Using these weights, we obtain a first estimate $\tilde{\beta}(y)$ by maximizing the weighted likelihood. Then we obtain new weights using this new estimate, and we iterate until convergence.

Let $t_1 < t_2 < ... < t_M \leq y_U$ be all observed failure times below y_U. The empirical CDR coefficient process is a right-continuous step function that jumps only at these values. The estimated CDF of Y^* is zero for all $y < t_1$.[3] Algorithm 1 provides the detailed steps for the discrete case. In the continuous case, we can skip the while-loop in lines 6 to 10.

Algorithm 1 CDR-MLE algorithm

1: Initialize $w(y, y_i, x_i) = 0$ for $i = 1, ..., n$.
2: Initialize $\tilde{\beta} = (-\infty, 0, 0, ..., 0)$.
3: **for** $j = 1, 2, ..., M$ **do**
4: Obtain $\hat{\beta}(t_j)$ by maximizing the log likelihood $log(L(t_j))$ in 2.3.3.
5: Update $w(y, y_i, x_i) = \frac{\Lambda(x_i'\hat{\beta}(t_i)) - \Lambda(x_i'\hat{\beta}(y_i))}{1 - \Lambda(x_i'\hat{\beta}(y_i))}$ for i such that $y_i < t_j$ and $\delta_i = 0$.
6: **while** $\|\tilde{\beta} - \hat{\beta}(t_j)\| > tol$ **do**
7: $\tilde{\beta} = \hat{\beta}(t_j)$
8: Obtain $\hat{\beta}(t_j)$ by maximizing the log likelihood $log(L(t_j))$ in 2.3.3.
9: Update $w(y, y_i, x_i) = \frac{\Lambda(x_i'\hat{\beta}(t_i)) - \Lambda(x_i'\hat{\beta}(y_i))}{1 - \Lambda(x_i'\hat{\beta}(y_i))}$ for i such that $y_i < t_j$ and $\delta_i = 0$.
10: **end while**
11: $\hat{\beta}_{MLE}(t_j) = \hat{\beta}(t_j)$
12: **end for**

This algorithm can be seen as a distribution regression analog of the quantile regression estimator of Portnoy (2003). He exploits the strategy to redistribute the weights of the censored observations to two pseudo-

[3]Equivalently, we can drop the observations censored below t_1 because they do not contain any information of the conditional distribution of Y^*.

observations. As discussed in Section 2.2, quantile regression is not well suited for discrete outcomes. Thus, Portnoy (2003) considers only this case and does not need to iterate until convergence. Our method for discrete outcomes also bears some similarities with the Buckley and James (1979) approach to censored linear regression. They have a similar problem because the correction depends on the true coefficients of the regression. They also solve this problem iteratively under convergence.

2.3.2 Martingale-based Approach

Peng and Huang (2007) consider the same model as we do but only for continuous durations. They propose to utilize the martingale structure of randomly right-censored data to construct an estimator. Since one of the advantages of DR (e.g., compared to quantile regression) consists of dealing with discrete data, we extend their framework and propose a new estimator that accommodates continuous, discrete, and mixed outcomes.

Define $N_i(y) = 1(Y_i \leq y, \delta_i = 1)$, $T_i(y) = Y_i \geq y)$, and

$$M_i(y) = N_i(y) - \int_0^y T_i(s) \frac{dF_{Y^*|X}(s|x_i)}{1 - F_{Y^*|X}(s-|x_i)} \qquad (2.3.4)$$

where the integral is defined as a Stieltjes integral regardless of whether $F_{Y^*|X}$ is absolutely continuous or not. $F_{Y^*|X}(s-|x_i)$ denotes the left-hand limit of $F_{Y^*|X}(s|x_i)$. For discrete distributions, $dF_{Y^*|X}(s|x_i) = F_{Y^*|X}(s|x_i) - F_{Y^*|X}(s-|x_i) = \mathrm{P}(Y^* = s|X = x_i)$. For continuous distributions, $dF_{Y^*|X}(s|x_i) = f_{Y^*|X}(s|x_i)ds$. The fraction corresponds to the conditional (discrete or continuous) hazard rate. $M_i(y)$ is a martingale process and has expectation zero conditional on X, see for instance Fleming and Harrington (2011) or Aalen et al. (2008).

If the outcome distribution is continuous and Assumption 5 is satisfied, we can write

$$M_i(y) = N_i(y) - \int_0^y T_i(s) \frac{d\Lambda(x'\beta(s))}{1 - \Lambda(x'\beta(s))}$$

In particular, if $\Lambda(x'\beta(s)) = 1 - exp(-exp(x'\beta(s)))$, it simplifies to

$$M_i(y) = N_i(y) - \int_0^y T_i(s) d\exp(x'\beta(s)) \qquad (2.3.5)$$

Peng and Huang (2007) suggest an estimator that is the sample analog of 2.3.5. Instead, we suggest an estimator base on 2.3.4. The difference is minimal for continuous outcomes, but this modification is necessary to accommodate mass-points. Interestingly, our modified estimator is numerically equal to the Kaplan-Meier estimator, which is not the case for the Peng and Huang (2007) estimator.

Remember that $t_1 < t_2 < ... < t_M \leq y_U$ are all observed failure times below y_U. Let $x_{(1)}, x_{(2)}, ..., x_{(M)}$ be the associated covariate vectors. In the presence of mass points, we take the sum of the covariates with an outcome equal to the corresponding value, i.e. $x_{(m)} = \sum_{i:y_i=t_m} x_i \cdot \delta_i$. We estimate the CDR process sequentially:

Algorithm 2 CDR-Martingale algorithm

1: Obtain $\hat{\beta}(t_1)$ as the value of b that solves the equation

$$x_{(1)} - \sum_{i=1}^{n} x_i T_i(t_1)\Lambda(x_i'b) = 0$$

2: **for** j=2,...,M **do**

3: Obtain $\hat{\beta}(t_j)$ as the value of b that solves the equation

$$x_{(j)} - \sum_{i=1}^{n} x_i T_i(t_j) \frac{\Lambda(x_i'b) - \Lambda(x_i'\hat{\beta}(t_{j-1}))}{1 - \Lambda(x_i'\hat{\beta}(t_{j-1}))} = 0$$

4: **end for**

2.4 Asymptotic Results

Peng and Huang (2007) derive the large-sample properties of their martingale estimator when the dependent variable is continuous. A similar proof strategy can be used to derive similar results for our modified martingale estimator. Note that the bootstrap should be recommended in both cases because the analytical estimation of the asymptotic variance is generally difficult.

We now study the large-sample behavior of the CDR-MLE estimator.

Assumption 8. *The vector X is uniformly bounded:* $\sup_i \|x_i\| < \infty$

Assumption 9. *The link function* $\Lambda : \mathcal{R} \mapsto (0,1)$ *is invertible, twice continuously differentiable, and its first derivative* $\lambda(\cdot)$ *is bounded away from zero.*

Assumption 10. *The minimum eigenvalues of*

$$E\left[\frac{\lambda(x_i'\beta(y))^2}{\Lambda(x_i'\beta(y))(1-\Lambda(x_i'\beta(y)))}1(y_i \geq y)x_i x_i'\right]$$

is bounded away from zero uniformly over $y \in (0, y_U]$.

Assumption 11. *Either the support of Y^* consists of a finite subset of $\mathbb{R}$ or it is a compact interval in $\mathbb{R}$. In the later case, the conditional density $f_{Y^*|X}(y|x)$ exists, is uniformly bounded away from zero and from above, and is uniformly continuous in (y, x) in the support of (Y^*, X).*

Assumption 8 is almost always met in practice. Assumption 9 puts additional restrictions on the link function. It is satisfied by all traditional link functions, e.g. by the complementary log log, logit, and probit link functions. Assumption 10 is an identification condition that imposes invertibility of the first-order derivative of the estimating equation. More intuitively, it excludes linearly dependent regressors among the surviving population. Assumption 8 imposes some regularity condition on the conditional density function when the outcome variable is continuous.

For the moment we have been able to derive the asymptotic distribution of the DR-MLE only with a stronger assumption to initialize the algorithm. To deal with the bottom of the distribution, we assume that there is no censoring below some threshold y_L:

Assumption 12. *When the outcome is continuous, there exists $y_L > 0$ such that $Pr(\delta_i = 1|y_i^* \leq y_L) = 1$.*

We need this assumption only for continuous outcomes. Portnoy (2003) makes a similar assumption for quantile regression. It is probably possible to avoid this assumption by adapting the proof strategy in Peng and Huang (2007) or Peng (2012).

Theorem 3. *Under Assumptions 1-10, as $n \to \infty$, the CDR coefficient process, $\sqrt{n}(\hat{\beta}_{MLE}(y) - \beta(y))$ weakly converges to a mean-zero Gaussian process for $y \in [y_L, y_U]$.*

Proof sketch. Consider first the range $y \in (0, y_L)$. Note that without censoring the estimator simplifies to the distribution regression estimator studied in Chernozhukov et al. (2013). Our assumptions imply their assumptions when $\mathrm{P}(\delta = 1|Y \leq y) = 1$. It follows by their Corollary 5.3 that the CDR coefficient process converges to a mean-zero Gaussian process on $y \in (0, y_L)$.

For the remaining threshold, our estimator maximizes the objective function in (2.3.3), which is the MLE of a weighted binary regression. If we knew the weights, the results in Chernozhukov et al. (2013) could be trivially extended. The issue is that we have estimated these weights in a previous regression such that the asymptotic distribution will involve a product integral of the process.

The first-order derivative of the likelihood function in (2.3.3) can be expressed as

$$\begin{aligned}\varphi_{y,\beta}(Y,X,\delta) = & \frac{\delta \cdot 1(Y \le y)}{\Lambda(X'\beta)}\lambda(X'\beta)X \\ & - \frac{\delta \cdot 1(Y > y)}{1-\Lambda(X'\beta)}\lambda(X'\beta)X \\ & - \frac{(1-\delta)\cdot 1(Y \ge y)}{1-\Lambda(X'\beta)}\lambda(X'\beta)X \\ & + (1-\delta)1(Y<y)\frac{\Lambda(X'\beta)-\Lambda(X'\hat{\beta}_{MLE}(Y))}{1-\Lambda(X'\hat{\beta}_{MLE}(Y))}\frac{\lambda(X'\beta)X}{\Lambda(X'\beta)} \\ & - (1-\delta)1(Y<y)\frac{1-\Lambda(X'\beta)}{1-\Lambda(X'\hat{\beta}_{MLE}(Y))}\frac{\lambda(X'\beta)X}{1-\Lambda(X'\beta)}\end{aligned}$$

The CDR-MLE estimator is the Z-estimator that satisfies $\|\widehat{\Psi}(\hat{\beta}_{MLE}(y),y)\| = 0$, where $\widehat{\Psi}(\beta,y) = \frac{1}{n}\sum_{i=1}^{n}\varphi_{y,\beta}(y_i,x_i,\delta_i)$. Note that the population analog is $\Psi(\beta,y) = E[\varphi_{y,\beta}(y_i,x_i,\delta_i)] = 0$. Under the assumptions made, the class of functions $\varphi_{y,\beta}(Y,X,\delta)$ is Donsker: the class of indicator variables is Donsker, $\lambda(\cdot)$ is bounded and differentiable, $0 < \Lambda(X'\beta) < 1$ and X is bounded.[4]

[4] It is in principle possible that the denominator $1-\Lambda(X'\beta)$ is equal to zero. However, in this case, both numerators $\Lambda(X'\beta)-\Lambda(X'\beta(Y))$ and $1-\Lambda(X'\beta)$ as well as $1(Y>y)$ are also equal to zero. Thus, the denominator can take a value of zero only when the whole term does matter. To simplify the notation we define $\frac{0}{0}=0$.

The derivative of $\Psi(\beta, y)$ with respect β is given by

$$
\begin{aligned}
J(y) = E\Big[&\lambda(x_i'\beta(y))^2 x_i x_i' \cdot \\
&\Big(\frac{\delta_i \cdot 1(y_i \leq y)}{\Lambda(x_i'\beta(y))^2} \\
&+ \frac{\delta_i \cdot 1(y_i > y)}{(1-\Lambda(x_i'\beta(y)))^2} \\
&+ \frac{(1-\delta_i)\cdot 1(y_i \geq y)}{(1-\Lambda(x_i'\beta(y)))^2} \\
&+ \frac{(1-\delta_i)\cdot 1(y_i < y)}{\Lambda(x_i'\beta(y))^2} \frac{\Lambda(x_i'\beta(y)) - \Lambda(x_i'\beta(y_i))}{1-\Lambda(x_i'\beta(y_i))} \\
&+ \frac{(1-\delta_i)\cdot 1(y_i < y)}{(1-\Lambda(x_i'\beta))^2} \frac{1-\Lambda(x_i'\beta(y))}{1-\Lambda(x_i'\beta(y_i))} \\
&+ \frac{(1-\delta_i)1(y_i < y)}{(1-\Lambda(x_i'\beta(y_i)))\Lambda(x_i'\beta)} \\
&+ \frac{(1-\delta_i)1(y_i < y)}{(1-\Lambda(x_i'\beta(y_i)))(1-\Lambda(x_i'\beta))}\Big)\Big]
\end{aligned}
$$

Note that $J(y)$ is invertible by Assumption 10.

$\Psi(\beta, y)$ is a function $\hat{\beta}_{MLE}(\tilde{y})$ for $\tilde{y} < y$. The derivative of $\Psi(\beta, y)$ with respect to $\hat{\beta}_{MLE}(\tilde{y})$ is

$$
\begin{aligned}
\Delta(y, y_i) \equiv E\Big[&\lambda(x_i'\beta(y))^2 x_i x_i' \cdot \\
&\Big(-\frac{(1-\delta_i)1(y_i < y)}{(1-\Lambda(x_i'\beta(\tilde{y})))\Lambda(x_i\beta(y))} \\
&+ (1-\delta_i)1(y_i < y)\frac{\Lambda(x_i'\beta(y)) - \Lambda(x_i'\beta(\tilde{y}))}{(1-\Lambda(x_i\beta(\tilde{y})))^2\Lambda(x_i'\beta(y))} \\
&- (1-\delta_i)1(y_i < y)\frac{1-\Lambda(x_i'\beta(y))}{(1-\Lambda(x_i'\beta(\tilde{y})))^2(1-\Lambda(x_i\beta(y)))}\Big)\Big]
\end{aligned}
$$

For $\tilde{y} < y_L$, by corollary 5.3 in Chernozhukov et al. (2013), $\hat{\beta}_{MLE}(\tilde{y})$ admits asymptotically the linear representation

$$
\sqrt{n}(\hat{\beta}_{MLE}(\tilde{y}) - \beta(\tilde{y})) = -J(\tilde{y})^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varphi_{y,\beta(y)}(y_i, x_i, \delta_i) + o_p(1)
$$

For large values of $\tilde{y}$, we apply sequentially Lemma E.1 in Chernozhukov et al. (2013) and obtain

$$\sqrt{n}(\hat{\beta}_{MLE}(y) - \beta(y)) =$$
$$- J(y)^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\varphi_{y,\beta(y)}(y_i, x_i, \delta_i) + \int_0^y \Delta_{y,\tilde{y}}\varphi_{\tilde{y},\beta(\tilde{y})}(y_i, x_i, \delta_i)d\tilde{y}\right) + o_p(1)$$

The result of the theorem follows.

■

2.5 Simulations

Subsequently, we address the finite sample performance of the CDR estimators introduced in the preceding sections. As outlined earlier, we expect that the two CDR estimators will provide asymptotically equivalent results. We present the results based on both approaches to assess whether this holds across different data generating processes (DGP) and sample sizes. Our focus lies on the performance across three different DGPs, which differ regarding a single regressor's effect. In particular, by imposing various degrees of effect heterogeneity, we will address the bias-variance trade-off when allowing for time-varying coefficients. For this purpose, we will compare the CDR estimators to the Cox model that serves as benchmark case in the following.

In the following, we will directly model the regression coefficients $\beta(y)$. The underlying CDF will have the form of $F_{Y|X}(y|X) = 1 - \exp(-\exp(X'\beta(y))$ where X includes a constant and potentially many regressors. We generate the sample distribution of Y_i by drawing a uniform random variable, u_i, for each observation and setting Y_i to the lowest value at which $F_{Y|X}(y|X) < u_i$. Prespecifying $\beta(y)$ allows us to govern the outcome distribution's shape without assuming a functional form of the baseline hazard. To obtain our benchmark results, it suffices to take the simple case in which X only includes a constant and a uniformly distributed regressor x_u. To address how time-varying effects affect performance, we will vary the effect of x_u: (i) $\beta_{x_u}(y) = 0$ for all y, (ii) $\beta_{x_u}(y) = -1 + 2/T \cdot y$ and (iii) $\beta_{x_u}(y) = -2 + 4/T \cdot y$ where T denotes the maximum value of the outcome. The effect of the constant will vary only marginally to ensure an equal fraction of censored observations across all DGPs. Figure 2.5 in the Appendix illustrates how the outcome and the corresponding censoring variable are distributed. We expect all models to provide unbiased estimates in case (i). Yet, as the true coefficients become more heterogeneous, we will necessarily observe a bias in the Cox models results'. On the other hand, we expect the CDR estimators to be unbiased in all DGPs but at the price of a higher variance. Our primary goal will be to assess the extent of this classical trade-off.

For all simulations, a fraction of 25% of the observations is censored. The censoring distribution is allowed to depend on x_u. This corresponds to the standard assumption regarding censoring, i.e. conditional independence. Further, T, the maximum duration, is set to 20. The results do not depend on the level of T; a simulation with T=100 can be found the in table 2.3 Appendix. The results of this section are based on four different sample sizes $n = \{250, 1000, 4000, 16000\}$ with 1000 Monte Carlo experiments each. Finally, note that all results are obtained using 48 cores and the `parallel` package in `R`.

First, we evaluate the estimated conditional CDF (CCDF). For this purpose, we define a grid of quantile values for x_u and Y at which the CCDF is evaluated. The grid consists of 25 equidistant points between the 10th and 90th quantile of x_u and Y, respectively. To obtain a conclusive measure of performance, we average the mean squared error (MSE), the standard deviation (SD), and the absolute bias across all cells specified by the grid. Table 2.1 reports the resulting statistics multiplied by 100. Let us consider the three DGPs separately. In the benchmark case of DGP 1, where x_u does not affect Y, the Cox model performs best. The CDR estimators obtain a slightly higher MSE due to a more significant dispersion reflected by the standard deviation. Expectedly, all models get unbiased estimates. Comparing the magnitude of the standard deviation and the bias, we observe that the former is much more important. Note that for all DGPs, the two CDR estimators provide almost identical estimates. Some deviances do occur at the very tails of the distribution. In the case of the present DGPs, the CDR-Martingale estimator seems to perform marginally better. Yet, the differences remain minimal indicating that the two approaches are asymptotically close.

Next, we consider the DGPs incorporating heterogeneous coefficients. As with DGP 1, the CDR model's standard deviation is about 10% higher than the one in the Cox model for all models. In contrast, the Cox model's estimates are biased once we allow for time-varying effects. Trading off this biasedness against the CDR model's modestly higher standard deviation, the results suggest that we prefer to have an unbiased estimate. For instance, based on DGP 2, the CDR-MLE estimator's standard deviation is roughly 11% larger than with the Cox model (n=4000). In contrast, the Cox models' bias is 98% higher, which increases the mean squared error by a factor of five. In relative terms, the bias accounts for 85% of the Cox models MSE in the same case. This underscores the higher relevance of the bias too. For

Table 2.1: Average MSE, Bias and SD of the CCDF (multiplied by 100)

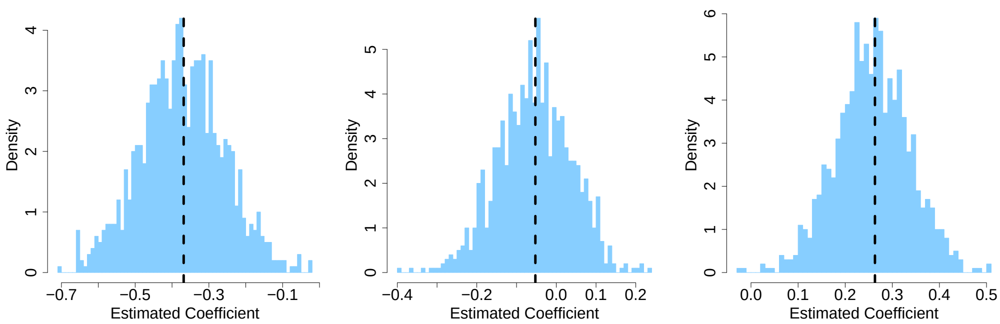
	Cox			CDR-MLE			CDR-Martingale		
n	MSE	SD	Bias	MSE	SD	Bias	MSE	SD	Bias
				DGP 1: $\beta_{x_u}(y) = 0$					
250	0.1171	3.3545	0.0904	0.1529	3.8064	0.1102	0.1559	3.8325	0.1082
1000	0.0310	1.7286	0.0518	0.0394	1.9374	0.0527	0.0393	1.9352	0.0528
4000	0.0078	0.8659	0.0234	0.0100	0.9737	0.0298	0.0100	0.9734	0.0296
16000	0.0020	0.4407	0.0087	0.0025	0.4907	0.0117	0.0025	0.4907	0.0119
				DGP 2: $\beta_{x_u}(y) = -1 + (2/T) \cdot y$					
250	0.1685	3.4058	1.7131	0.1540	3.8339	0.0979	0.1538	3.8324	0.0885
1000	0.0796	1.7309	1.7191	0.0388	1.9256	0.0559	0.0387	1.9229	0.0530
4000	0.0560	0.8706	1.7132	0.0099	0.9701	0.0298	0.0099	0.9689	0.0306
16000	0.0507	0.4407	1.7218	0.0025	0.4864	0.0094	0.0025	0.4862	0.0093
				DGP 3: $\beta_{x_u}(y) = -2 + (4/T) \cdot y$					
250	0.2370	3.3760	2.6846	0.1544	3.8302	0.1192	0.1549	3.8336	0.1084
1000	0.1520	1.7198	2.7263	0.0386	1.9203	0.0435	0.0384	1.9154	0.0399
4000	0.1289	0.8684	2.7226	0.0100	0.9727	0.0303	0.0099	0.9700	0.0317
16000	0.1240	0.4377	2.7340	0.0025	0.4861	0.0090	0.0025	0.4855	0.0086

Notes: Throughout, a fraction of 25% of the observations were censored where censoring was allowed to be a function of the regressor. For the outcome as well as for the regressor, a grid of 25 quantiles ranging from .1 to .9 has been used. For each of these 25^2 *cells, the MSE, bias and variance were computed separately. The table represents the averages over all cells for 1000 replications. In the case of time-varying coefficients* $\beta_{x_u}(y)$ *has been chosen such its average over T roughly equals* β_{x_u} *from DGP 1.*

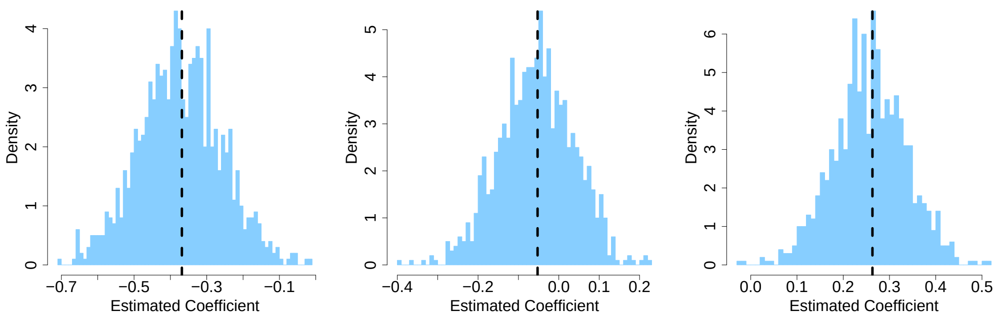
small data sizes, the difference between the MSE of the two models remains small. The reason herefore is that in these cases, the standard deviation accounts for most of the MSE. Based on table 2.1, we conclude that the CDR estimator's unbiasedness comes at a relatively low cost in the form of a higher standard deviation. Thus, in a case where we suspect time-varying effects to be prevalent, we prefer to use the CDR estimator.

Note that by construction, deviations from the underlying CDF are lower at the tails of the distribution. To account for this, table 2.2 in the Appendix reports a corrected version of the statistics from table 2.1. In particular, we divide the MSE, standard deviation, and bias by $F_{Y|X}(y|X)(1 - F_{Y|X}(y|X))$, i.e., the variance at the respective point of the grid. Note that this correction is done before averaging across all entries of the grid. In terms of the bias-variance trade-off, the results in table 2.1 suggest that the above insights remain unchanged. The flexibility of the CDR estimators pays off as, in relative terms, the difference between the two models becomes slightly more substantial. More precisely, the CDR-MLE estimator reduces the Cox models MSE by 51% without the correction and by 55% using the correction (DGP 2, n = 4000). This indicates that the CDR estimator outperforms the Cox model by an even larger margin at the tails of the distribution.

Figure 2.1: Distribution of the $\hat{\beta}_{x_u}(y)$ for CDR-MLE and CDR-Martingale



(a) MLE, $Y = Q_Y(.25)$ (b) MLE, $Y = Q_Y(.5)$ (c) MLE, $Y = Q_Y(.75)$

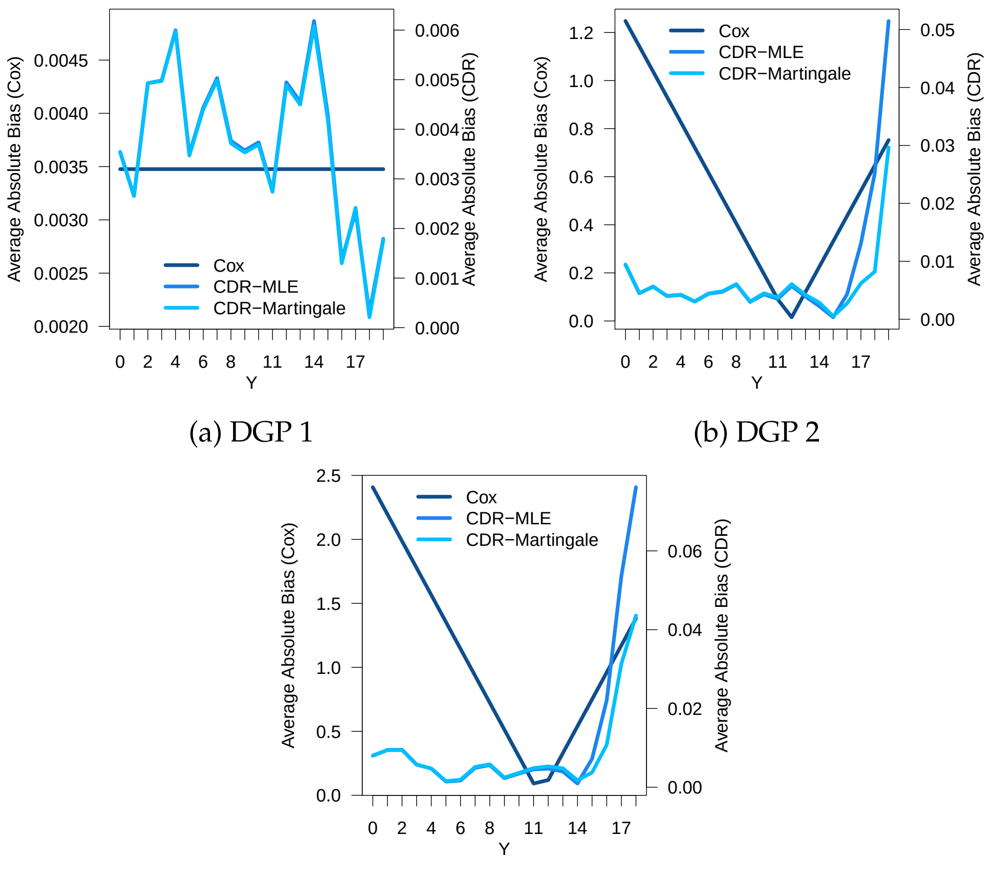


(d) Mart., $Y = Q_Y(.25)$ (e) Mart., $Y = Q_Y(.5)$ (f) Mart., $Y = Q_Y(.75)$

Notes: The figure presents the estimated coefficients based on 1000 draws from DGP 2 with $n = 4000$ and $T = 20$. The dashed black line marks the true values of $\beta_{x_u}(y)$ at the respective value of Y.

In the next step, we assess the distribution of the estimated coefficients. As suggested above, we expect the CDR models' regression coefficients to be asymptotically normal. In figure 2.1, we present the estimated coefficients at the 25th, 50th, and 75th quantile of Y for the CDR-MLE and the CDR-Martingale estimator. All panels of figure 2.1 suggest that the coefficients are roughly normally distributed around the true value, which is in line with the asymptotic distribution. Further, the two estimators seem to match closely in terms of the distribution of $\hat{\beta}_{x_u}(y)$. In particular, the variance and shape of the distribution are almost identical. Taken together, figure 2.1 provides further evidence that the two estimators are closely related.

Figure 2.2: Bias of $\hat{\beta}_{x_u}(y)$ as a Function of Y



(a) DGP 1

(b) DGP 2

(c) DGP 3

Notes: The figure illustrates the average absolute bias multiplied by 100 for the CDR-MLE, CDR-Martingale and the Cox estimator. The data size is kept constant at $n = 4000$.

Finally, let us consider the estimated coefficients across the distribution of the outcome. Considering that in our DGPs, the regression coefficients are a linear function of time, we expect that the Cox model is biased strongest at the tails. Figure 2.2 describes the average absolute bias multiplied by 100 as a function of the outcome. The three panels of Figure 2.2 refer to the three different DGPs. In the benchmark case incorporating time-constant effects, all estimators provide unbiased estimates. As a consequence, the absolute biases are close to zero. The results based on the two remaining DGPs suggest that the Cox model's bias is largest at the bottom of the distribution. Further, the Cox model's bias is a linear function of time because the true coefficients increase linearly. As the Cox model coefficients represent a weighted average of the true time-varying effects, the bias decreases until the true coefficients equal the estimated ones. Then, again, the bias rises. Considering the two CDR estimators, we observe that they differ only at the top of the outcome's distribution. The CDR-Martingale estimator performs slightly better, which is in line with our earlier findings from table 2.1. The difference between the CDR estimators is minimal, again confirming our expectation that the two approaches are asymptotically equivalent.

2.6 Application to Unemployment Duration

The optimal design of unemployment insurance (UI) systems has been a focus of economic research because of its relevance for welfare. For instance, Lalive et al. (2015) find that UI policies affect equilibrium outcomes for all market participants, not only the workers eligible to receive UI benefits. Among the many instruments available to policy-makers, setting the potential benefit duration (PBD) is a frequently used tool to guard incentives. Typically, the PBD depends on age and on how long individuals have been paying insurance premiums. The present application analyzes how an increase in PBD affects unemployment duration. In particular, we will focus on whether the effect varies over the unemployment spell.

From a microeconomic perspective, optimal UI faces a well-known insurance-incentive trade-off involving moral hazard (Chetty, 2006). More precisely, UI benefits (i) allow for consumption smoothing (ii) but reduce job searching incentives. The former may be especially important for liquidity-constrained households as without financial support, these individuals cannot smooth consumption (Chetty, 2008; Landais, 2015). As a consequence, individuals are heterogeneously affected by UI policies. Apart from differences across individuals, there is a second type of heterogeneity involved: Typically, factors determining the probability of finding a new job have time-varying effects. For instance, numerous studies have found that job search behavior depends on reservation wages, the generosity of benefits, or individual beliefs. As suggested by Krueger and Mueller (2016), reservation wages decline throughout unemployment, increasing the likelihood of accepting job offers. Marinescu and Skandalis (2021) find that job seekers intensify their attempts to find a job prior to benefit exhaustion which supports that individuals' efforts are too low because of moral hazard. Finally, Mueller et al. (2021) provide evidence showing that beliefs about the likelihood of being reemployed are subject to an optimistic bias. In particular, the bias is more substantial for long-term unemployed and not corrected downward over time. According to this finding, individuals soon reaching the PBD may suddenly realize that they will no longer get benefits. Taken together, we expect that, on average, an increase in PBD leads to higher unemployment spells. Further, we suspect that closer to the exhaustion of PBD, individuals are more likely to find a job as they adapt their target wages and intensify their efforts. Consequentially, being granted a longer PBD would have the most substantial effect at the very top of the durations' distribution.

In the following, we analyze the distributional effects of an increase in PBD in Switzerland. We exploit that above 24 years of age, individuals are granted a PBD of 18 instead of 9 months.[5, 6, 7] Adopting a similar discontinuity at an age threshold, Lalive (2007) finds that a longer PBD of 170 weeks increases effective duration by 15 weeks for men.[8] As suggested by Johnston and Mas (2018), the same logic applies to cuts in unemployment durations too. The authors analyze a cut introduced by a new law and find that a 1-month reduction of maximal duration reduces the unemployment spell by 1.8 weeks. Evidence on the distributional effects of longer PBD is scarce. Melly and Lalive (2020) study the quantile treatment effects of longer PBD based on data from Lalive (2007) and find positive and increasing effects on duration. Yet, in their case, censoring is not an issue

[5]The law introducing this cut-off entered into force as from the 1st of April 2011 (see *Bundesgesetz über die obligatorische Arbeitslosenversicherung und die Insolvenzentschädigung*). We will consider data for unemployment spells starting after this date only.

[6]If individuals turn 25 during unemployment, they automatically switch regime and are granted a PBD of 18 months. As a consequence, the age at the beginning of unemployment is decisive: individuals younger (older) than 24 years and three months get a PBD of 9 months (18 months).

[7]Note that the PBD is set to 4 months for individuals that have not been paying contributions into the UI for more than one year. As this holds irrespective of age, it does not affect the estimation based on the cut-off at 24 years and three months.

[8]The same effect is much stronger for women resulting in a longer duration of 110 weeks. This effect is partly due to another incentive: to enter a favorable social security scheme later on.

which enables the use of standard distributional methods. Relatedly but studying a different effect, Delgado et al. (2019) examine the diversity of UI benefits' effect on unemployment duration. The results indicate that higher benefits have a U-shaped impact, i.e., the benefits increase duration most for average durations. Further, taking net wealth as a proxy for liquidity, the authors argue that the effects are more substantial for individuals facing liquidity constraints.

We use cross-sectional data from the Swiss State Secretariat for Economic Affairs (SECO) on all individuals registered at a regional employment agency irrespective of whether they receive benefits. We consider individuals that start being unemployed after the 1st of April 2011 and are no longer unemployed on the 1st of May 2018. Further, we restrict our estimation sample to include only singles at the age of 19 to 28 at the start of unemployment. The resulting estimation sample captures 437,909 observations. Note that the observed UI duration is right-censored as some individuals (i) renounce from help and UI benefits (14 % of the data), (ii) do not obey the rules of the regional employment agency (10%), (iii) go abroad (1%) or drop out because of other unkown reasons (14%).

To assess how PBD effects unemployment spells, we will estimate the model in equation (2.6.1) using the CDR-MLE estimator. We model the conditional distribution of unemployment spells as

$$F_{Y|X}(y|x) = \Lambda(x'\beta(y)), \tag{2.6.1}$$

where $\Lambda(\cdot)$ denotes the complementary log-log link function, y is the duration of unemployment, and X includes a constant and the following covariates: a treatment indicator for a being older than 24 at the start of unemployment, a linear and quadratic term for age at the start, gender, nationality, education, experience and whether the mother tongue is German.[9] All of these regressors are predetermined and can therefore not change during unemployment. Our estimated effects have a causal interpretation if individuals below and above the threshold are comparable. As we include a linear and quadratic term for age at the start of unemployment, we account for average differences between younger and older individuals. Table 2.4 in the Appendix compares individuals above and below the threshold. The descriptive statistics suggest that the two groups are comparable, which supports our identification strategy.

[9]Nationality included as a dummy variable for being Swiss; education is grouped in three categories: mandatory school only, secondary school or tertiary education; and experience is measured in four categories: no, less than one year, one to three years or more" than three years of experience.

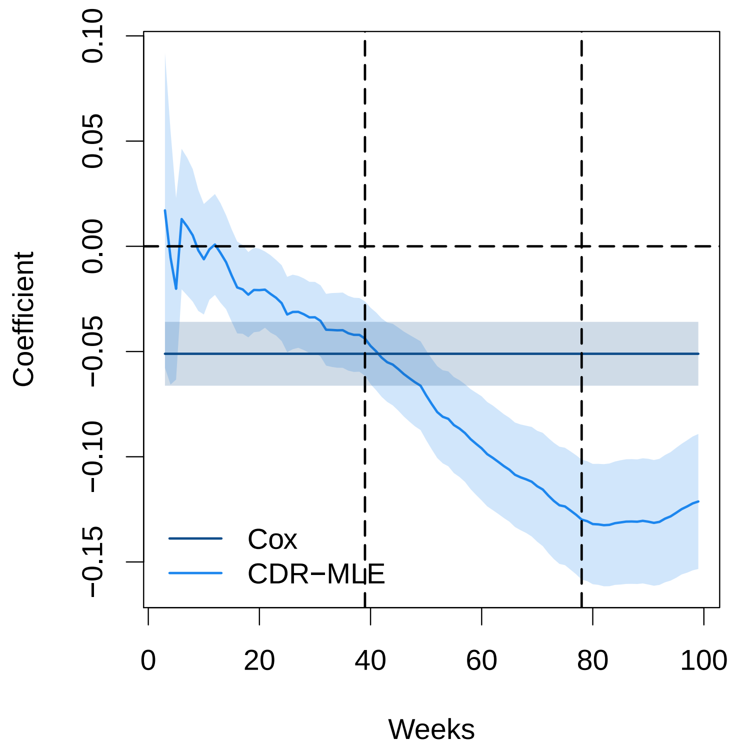
Considering the results, we first focus on the coefficient of our treatment variable. As the CDR-MLE estimator is a strict generalization of the proportional hazards model, we will present both models' estimated coefficients. This allows us to address the heterogeneity of the treatment effect directly. The coefficients estimated by CDR are difficult to interpret in the presence of time-varying effects. To ease interpretation, we complement our analysis by addressing the treatmentâ€™s unconditional effect too. Following the approach of Chernozhukov et al. (2013), we can recover the full distribution of unemployment spells. In particular, we compute the distribution of spells that only depends on our treatment variable, abstracting from all covariates' effects. We first compute the unconditional distribution of unemployment spells for both the treated and non-treated, $F_{Y\langle D=0\rangle}(y)$ and $F_{Y\langle D=1\rangle}(y)$, where $D = 1$ indicates being older than 24. This is done by first modifying the covariate distribution to have all entries of the treatment variable equal to zero or one, $X_{D=0}$ and $X_{D=1}$, respectively. Next, we integrate over the covariates' distribution according to equation (2.6.2).

$$F_{Y\langle D\rangle}(y) = \int_{\mathcal{X}} F_{Y|X}(y|x) dF_{X_D}(x). \tag{2.6.2}$$

Intuitively, the resulting counterfactual $F_{Y\langle D\rangle}(y)$ can be interpreted as the distribution of spells if everyone was treated (non-treated). Most importantly, note that $F_{Y\langle D\rangle}(y)$ no longer depends on the control variables. Taking the left inverse of $F_{Y\langle D\rangle}(y)$, we obtain the quantile functions for both the treated and non-treated. Finally, we define the quantile treatment effect (QTE) as the difference between the two quantile functions, i.e., $\Delta^{QE}(\tau) = Q_{Y\langle D=1\rangle}(\tau) - Q_{Y\langle D=0\rangle}(\tau)$.

Figure 2.3 reports the regression coefficients for both the Cox model and the CDR model. Longer durations imply a lower hazard, reflected in a negative regression coefficient. Thus, we expect to observe a negative coefficient throughout the distribution. The vertical, dashed lines represent the two different values of PBD: 39 and 78 for weeks for individuals below or above the threshold, respectively. The Cox estimates suggests that being above the threshold when laid off increases unemployment duration on average. More precisely, the results indicate that the treatment decreases the likelihood of getting a job by roughly 5% ($\exp(-.05) = .95$, i.e., the hazard is reduced by 5%). The estimates of the CDR-MLE model point towards the same direction. Yet, we observe an evident and negative trend once spells are higher. At the bottom of the distribution, PBD does not affect the likelihood of finding a job. In contrast, the long-term unemployed experience significantly higher durations when treated. Accordingly, the long-term unemployed below the threshold are increasingly likely to find

Figure 2.3: Effect of longer PBD: Coefficients

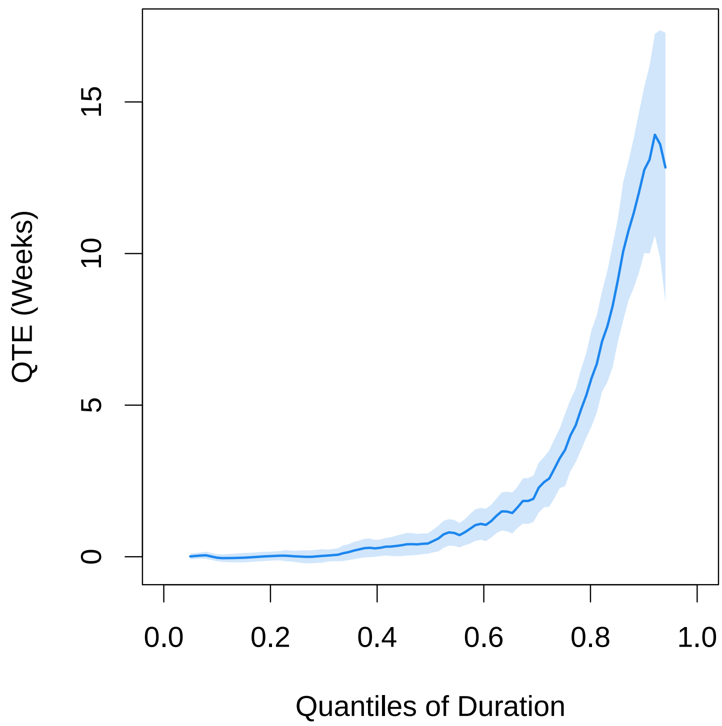


Notes: The vertical lines correspond to the PBD for the non-treated and treated, i.e., 39 and 78 weeks. The x-axis captures the distribution of unemployment durations from the 1st to the 99th quantile. All confidence bands are based point-wise standard errors.

a job once they are close to reaching their PBD. Based on the evidence mentioned above, this behavior is consistent with (i) intensified search, (ii) lower target wages, (iii) or short-sightedness if individuals had too optimistic beliefs.

Finally, we turn to the QTE presented in figure 2.4. The results suggest that for the lower half of the distribution, PBD has no relevant effect on duration. As we move upwards, we observe that the QTE rises to 14 weeks at the very top of the distribution. Thus, individuals in the top 20% of the distribution (spell $>$ 40 weeks) experience a substantial increase in unemployment duration if treated. This supports our previous findings from the regression coefficients only. We conclude that a higher PBD increases unemployment spells on average, with the effect being much more substantial for the long-term unemployed. Our finding contributes to a large body of studies focusing on the impact of UI policies. We leave it for future research to address the implications for welfare. For instance, incorporating time-varying effects into models like the ones in Landais (2015) or Lalive et al. (2015) may provide additional insights.

Figure 2.4: Effect of longer PBD: Quantile Treatment Effect



Notes: The x-axis captures the distribution of unemployment durations from the 5st to the 95th quantile. All confidence bands are based point-wise standard errors.

2.7 Conclusion

In this paper, we introduce an estimator (CDR-MLE) that deals with time-varying effects in the context of duration data. Our approach is based on distribution regression and represents a generalization of the famous model introduced by Cox (1972). Typically, handling censored observations is challenging in the case of heterogeneous effects. We overcome this difficulty by introducing pseudo-observations as weights when recursively estimating the conditional distribution. This estimation procedure applies to discrete and mixed outcomes as well as continuous ones - an attractive feature in applied work. Further, we allow for dependence between regressors and the censoring mechanism, i.e., the literature's standard assumption. In our main results, we establish consistency and asymptotic normality of the CDR-MLE estimator. In addition to these novel results, we highlight links to an existing estimator introduced by Peng and Huang (2007). In its original form, this estimator applies to continuous outcomes only. We extend their estimator to accommodate mixed and discrete outcomes. In this regard, we complement the literature by linking ideas and incorporating them into a unified framework.

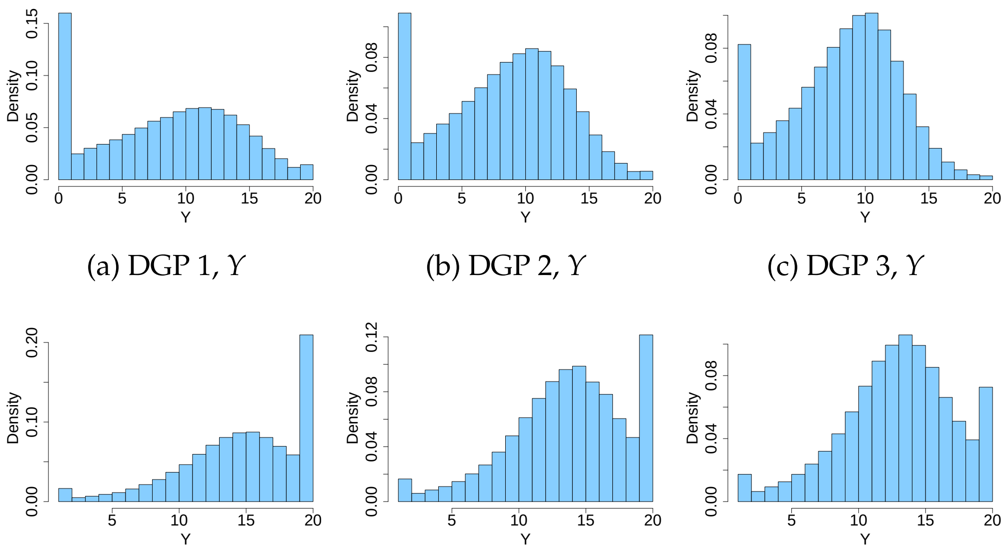
In the presence of heterogeneous effects, estimates based on Cox's proportional hazard model are biased. In contrast, the CDR-MLE estimator is unbiased at the cost of moderately higher variance. With respect to this trade-off, our simulation study results suggest that the bias dominates already for reasonable degrees of heterogeneity. In small samples, too, the CDR-MLE estimator achieves good finite sample performances. Further, we compare the CDR-MLE estimator's performance to the modified estimator based on Peng and Huang (2007). The results suggest that both CDR estimators are qualitatively equivalent.

We apply the CDR estimator to analyze how potential benefit duration (PBD) affects unemployment duration. Search models predict that individuals adapt their behavior over the unemployment spell, implying heterogeneous effects of PBD. Our results suggest PBD increases duration on average but much pronounced towards the top of the durations distribution. This result is in line with individuals intensifying their search efforts and adapting reservation wages once faced with the upcoming exhaustion of benefits. Thus, we can document the heterogeneity of an essential empirical effect likely to be missed by standard approaches using CDR.

The set-up of distribution regression would allow for numerous extensions. For instance, the CDR-MLE estimator could be adapted to capture the case of fixed effects or instrumental variables. Similarly, one could address the challenges of big data. To date, there is no formal result on the equivalence of the CDR-MLE and the CDR-Martingale estimator. Peng (2012) has obtained an analog result in the case of quantile regression. It seems to be a promising route for future research to address this task.

2.A Simulation Results

Figure 2.5: Outcome and Censoring Distribution for all DGP



(a) DGP 1, Y (b) DGP 2, Y (c) DGP 3, Y

(d) DGP 1, C (e) DGP 2, C (f) DGP 3, C

Notes: The figure presents the distribution of Y and the corresponding censoring distribution for all DGPs. The densities have been obtained using a sample of 100'000 observations.

Table 2.2: Corrected Average MSE, Bias and SD of the CCDF (multiplied by 100)

	Cox			CDR-MLE			CDR-Martingale		
n	MSE	SD	Bias	MSE	SD	Bias	MSE	SD	Bias
DGP 1: $\beta_{x_u}(y) = 0$									
250	0.6236	18.3887	0.6123	0.8457	21.2406	0.6930	0.8786	21.5161	0.6952
1000	0.1641	9.4450	0.3334	0.2169	10.7781	0.3020	0.2160	10.7603	0.3025
4000	0.0411	4.7152	0.1258	0.0546	5.3884	0.1610	0.0546	5.3860	0.1600
16000	0.0106	2.3989	0.0527	0.0138	2.7174	0.0742	0.0138	2.7169	0.0758
DGP 2: $\beta_{x_u}(y) = -1 + (2/T) \cdot y$									
250	0.9369	18.7320	10.4827	0.8385	21.4144	0.6861	0.8376	21.4081	0.5947
1000	0.4654	9.5042	10.5176	0.2092	10.6988	0.3143	0.2081	10.6729	0.2916
4000	0.3398	4.7732	10.4711	0.0531	5.3725	0.1636	0.0529	5.3621	0.1661
16000	0.3113	2.4142	10.5044	0.0133	2.6946	0.0614	0.0133	2.6926	0.0610
DGP 3: $\beta_{x_u}(y) = -2 + (4/T) \cdot y$									
250	1.3601	18.5639	16.5036	0.8363	21.4017	0.8074	0.8483	21.5098	0.7049
1000	0.9119	9.4650	16.6842	0.2087	10.7217	0.2515	0.2070	10.6789	0.2203
4000	0.7890	4.7708	16.6519	0.0534	5.4055	0.1724	0.0529	5.3824	0.1767
16000	0.7627	2.4049	16.6983	0.0133	2.7046	0.0621	0.0133	2.6975	0.0592

Notes: Throughout, a fraction of 25% of the observations were censored where censoring was allowed to be a function of the regressor. For the outcome as well as for the regressor, a grid of 25 quantiles ranging from .1 to .9 has been used. For each of these 25^2 cells, the MSE, bias and variance were computed separately. The table represents the averages over all cells for 1000 replications. In the case of time-varying coefficients $\beta_{x_u}(y)$ has been chosen such its average over T roughly equals β_{x_u} from DGP 1. All averages are divided by $F_{y|X}(y|X)(1 - F_{y|X}(y|X))$ to account for lower average errors at the tails.

Table 2.3: Continuous Case ($T = 100$): Average MSE, Bias and SD of the CCDF (multiplied by 100)

	Cox			CDR-MLE			CDR-Martingale		
n	MSE	SD	Bias	MSE	SD	Bias	MSE	SD	Bias
				DGP 1: $\beta_{x_u}(y) = 0$					
250	0.1206	3.4013	0.1078	0.1576	3.8633	0.1246	0.1646	3.9267	0.1293
1000	0.0314	1.7395	0.0548	0.0400	1.9516	0.0540	0.0399	1.9501	0.0543
4000	0.0079	0.8705	0.0220	0.0101	0.9793	0.0291	0.0101	0.9791	0.0290
16000	0.0020	0.4420	0.0076	0.0026	0.4930	0.0105	0.0026	0.4930	0.0106
				DGP 2: $\beta_{x_u}(y) = -1 + (2/T) \cdot y$					
250	0.1684	3.4027	1.7107	0.1550	3.8418	0.1002	0.1550	3.8413	0.0947
1000	0.0798	1.7379	1.7206	0.0392	1.9349	0.0681	0.0391	1.9327	0.0647
4000	0.0560	0.8711	1.7170	0.0099	0.9710	0.0293	0.0099	0.9699	0.0298
16000	0.0506	0.4408	1.7257	0.0025	0.4879	0.0090	0.0025	0.4876	0.0088
				DGP 3: $\beta_{x_u}(y) = -2 + (4/T) \cdot y$					
250	0.2430	3.4427	2.7160	0.1589	3.8929	0.1205	0.1591	3.8929	0.1140
1000	0.1546	1.7314	2.7409	0.0394	1.9375	0.0740	0.0393	1.9340	0.0700
4000	0.1312	0.8725	2.7432	0.0101	0.9790	0.0312	0.0101	0.9769	0.0321
16000	0.1264	0.4412	2.7536	0.0025	0.4906	0.0094	0.0025	0.4899	0.0090

Notes: Throughout, a fraction of 25% of the observations were censored where censoring was allowed to be a function of the regressor. For the outcome as well as for the regressor, a grid of 25 quantiles ranging from .1 to .9 has been used. For each of these 25^2 *cells, the MSE, bias and variance were computed separately. The table represents the averages over all cells for 1000 replications. In the case of time-varying coefficients* $\beta(t)$ *has been chosen such its average over T equals* β *from DGP 1.*

2.B Application Results

Table 2.4: Descriptive Statistics by Treatment and Control Group

	All	Control	Treated
Female	0.43	0.42	0.43
	(0.49)	(0.49)	(0.49)
Swiss	0.70	0.73	0.66
	(0.46)	(0.44)	(0.47)
German Speaking	0.44	0.46	0.41
	(0.50)	(0.50)	(0.49)
Experience <1Y	0.18	0.21	0.16
	(0.39)	(0.41)	(0.36)
Experience 1Y - 3Y	0.37	0.39	0.34
	(0.48)	(0.49)	(0.47)
Experience >3Y	0.35	0.27	0.44
	(0.48)	(0.44)	(0.50)
Secondary School	0.72	0.80	0.64
	(0.45)	(0.40)	(0.48)
Tertiary School	0.11	0.02	0.20
	(0.31)	(0.15)	(0.40)
Obs.	437909	233683	204226

Notes: Treatment and control group refers to individuals older (younger) than 24 years of age at the start of unemployment.

Chapter 3

Physician Induced Demand and Financial Incentives

Evidence from Large-Scale Fee Changes

joint with **Tamara Bischof**

Abstract

This paper analyzes how physicians adapt their provision of medical services when their financial incentives change. We exploit a plausibly exogenous and large-scale reform to the reimbursement system constituting a natural (quasi-)experiment. We find that physicians are not immune to monetary incentives. Conceptually, physicians may substitute across services or increase their aggregate supply in response to the new scheme. Our rich dataset allows us to isolate both channels carefully. We find that providers increase (decrease) the volume of services that became relatively more (less) attractive. The results vary considerably by the extent of the income loss. Further, physicians supply more services and treat more patients if they lose a substantial share of their revenue. We quantify the relative importance of the two response channels in terms of total healthcare costs.

Acknowledgements: We would like to thank Blaise Melly, Michael Gerfin and Kaspar Wüthrich for valuable comments and suggestions and Fabienne Töngi for the research assistance. We would also like to thank participants at the Brown Bag Seminar in Bern and the 12th World Congress of the Econometric Society for valuable comments. Moreover, we are grateful to Sasis AG (Oliver Grolimund and Sandra Wüthrich) for providing the data. We gratefully acknowledge financial support received from the Berne University Research Foundation (17/2018).

In the short run, the volume expansion channel is far more important than the aggregate effect of changes in substitution patterns. Finally, we evaluate the reform in light of the officially announced saving target and propose improvements for future policy interventions.

3.1 Introduction

The cost for the health sector has dramatically risen in recent years for the OECD countries. For instance, between 2007 and 2017, healthcare spending in terms of GDP increased from 12.4 to 17.1% for the US and 9.7 to 12.3% for Switzerland (OECD, 2019). Many experts call for cost-containment measures on the supply side to complement existing measures on the demand side (e.g. deductibles, co-payments). Most prominently, fee-for-service (FFS) is often thought to drive costs (McClellan, 2011, for a recent overview). Policymakers share this view: For instance, reforms of Medicare are based on the assumption that physicians offset half of their revenue losses due to fee reductions by increasing volume.[1] The empirical evidence on this topic, however, is mixed.

The purpose of this paper is to study how financial incentives affect physician behavior in terms of volume and type of medical services. The paper aims to disentangle and compare two response channels: the response due to changes in relative prices of medical services and the response due to changes in physician income.

We explore the effects of a large-scale fee change that is plausibly exogenous, by using rich physician-and-service level data from mandatory health insurance. In Switzerland, providers of ambulatory care are reimbursed under a nationally uniform fee-for-service (FFS) schedule called *TARMED*. In 2018, a reform of *TARMED* led to change in the fees which provides an ideal setting to analyze how financial incentives affect the supply of medical services. First, the regulator announced the new fees on relatively short notice, so the changes to the reimbursement system were arguably exogenous from the perspective of individual physicians. Second, in a fee-for-service system, the marginal procedure is fully reimbursed thus offering providers a financially rewarding response margin (as opposed to prospective payment systems). Third, fee changes have direct conse-

[1] The Health Care Financing Administration, commissioned to implement Medicare, assumes a fifty percent volume offset in their calculations, so that physicians are expected to increase their service volume after a fee reduction to halve their revenue loss. (cf. Codespote et al., 1998; Congressional Budget Office, 2007). See also the discussion in Reinhardt (1999).

quences on physician income because many physicians occupy a dual role of medical professional and entrepreneur (in 2014, 84% were (co-)owners; Hostettler and Kraft, 2018). Fourth, anecdotal evidence suggests that at least some physicians changed their treatment in response to the reform (swissinfo, 2018).

Using a flexible fixed effect regression approach, we identify the effect of fee changes on volumes of specific services and thus describe the way that financial incentives affect substitution across services (*substitution channel*). To the best of our knowledge, this paper provides the first attempt to address substitution patterns across the full set of medical services charged by a physician.

We find that fees affect how physicians substitute across services; namely, they increase volumes for services that became relatively more attractive, and reduce the supply of services with above-average fee cuts. Specifically, a fee increase by 4% above the physician-specific average raises the volume of the corresponding service by 1%. This behavior is consistent with profit-maximization.

Because the fee changes are unevenly distributed across medical specialties, we can additionally explore how the effect differs by the extent of the income loss. For example, radiologists lose a large chunk of their revenue (median of -13%), whereas other specialties are net winners. Indeed, the substitution pattern described above varies by how much revenue is at stake: This behavioral margin seems especially crucial for medical specialties that only experience moderate revenue losses or eventually benefit, such as general practitioners (GPs). In contrast, we find little evidence for substitution for doctors who lose a large fraction of their revenue (and thus, income).

Medical professionals may also react through a second channel; fee cuts translate to income losses, which may be (partly) recouped by providing more care. We label this the *volume expansion channel*. Service expansion may occur along two lines: treating more patients or increasing the service intensity per patient. We find evidence for such expansion, especially for providers that lose revenue. For each additionally lost revenue percent, physicians who lose revenue increase consultations by up to 2.6% and total service volume by roughly 0.5%. The volume expansion channel proofs especially relevant for the group that was found to have only a weak substitution response in the first analysis. These results suggest that for physicians who encounter a substantial income loss, increasing overall supply may be more important than substituting across services.

In addition to neatly separating the two response channels, we are able to quantify their relative importance.The main driver in terms of total healthcare costs is the volume expansion channel. A 5% revenue loss increases supply by roughly 3%. On the other hand, the substitution channel has little effect on costs. To some extent the negligible cost effect of substitution may be due to short-run restrictions on the physicians' service mix. The ability to address the relevance of the two channels allows us to draw valuable policy implications. Specifically, tying the fees to a service's value may leave healthcare costs unaffected in the short run, while possibly increasing quality.

Our findings tie to a body of literature studying how physicians react to incentives besides patient needs. For instance, previous research provides convincing evidence that obstetricians perform more C-sections when fertility declines (Gruber et al., 1999; Gruber and Owings, 1996), and that cardiac surgeons intensify treatment when they lose revenue (Yip, 1998). Our paper most closely relates to the empirical literature studying the effect of variation in fees (Brekke et al., 2017; Clemens and Gottlieb, 2014; Gruber et al., 1999; Rice, 1983; Yip, 1998).[2] Recently, Clemens and Gottlieb (2014), leveraging a large geographic consolidation in Medicare fees, find that a 2% increase in reimbursement rates leads to a 3% increase in healthcare provision. Similarly, we address the issue of expanding aggregate supply and contribute by extending the scope of our analysis to a wide set of services and closely examining the ensuing substitution patterns.

This line of research closely links to the literature on physician induced demand (PID), pioneered by Evans (1974). Under the PID hypothesis, physicians influence demand for medical services *against* the best interest of their patients (McGuire, 2000). McGuire and Pauly (1991) provide a widely used theoretical framework to study the PID phenomenon in Fee-for-Service (FFS) systems.[3] In their model, providers react to fee reductions by substituting supply towards more favorable services (substitution effect), and by increasing their overall healthcare supply (income effect, only present when physicians seek an income target). An empirical investigation of the strength of the income effect then provides evidence regarding the presence of PID versus profit-maximizing agents. The interpretation of

[2]The earlier empirical literature mostly focused on variations in the physician density (Cromwell and Mitchell, 1986; Fuchs, 1978; Stano, 1985, subject to critique by Dranove and Wehner, 1994; Gruber and Owings, 1996). Recently, similar approaches are used to study the effect of market concentration and competition (Dunn and Shapiro, 2018). More recently, there is a focus on the variation in the information asymmetry between medical professional and patient (Currie et al., 2011; Domenighetti et al., 1993; Gottschalk et al., 2020; Johnson and Rehavi, 2016; Schmid, 2015). See Chandra et al. (2011); McGuire (2000) for overviews.

[3]Other theoretical models are developed in Fuchs (1978); Zweifel et al. (1997).

our results is loosely based on the theoretical model, and our methodology is similar to Yip (1998), who aims to account for the substitution and the income effect in studying how the provision of cardiac surgeons react to a large fee change in Medicare. She finds that physicians with the highest income reduction increase the volume of an intensive procedure the most. By extending her work, we aim to isolate the two effects. Our methodology improves on her approach by using a more diverse population, a wider set of services, and a richer dataset.

More broadly, the paper relates to the literature on agency. As already pointed out by Arrow (1963), physicians act as agents for their patients, who delegate treatment decisions to their doctors. If physicians act as perfect agents, monetary incentives should play no role in determining treatment, as physicians assist their patients in demanding the optimal quantity of services. However, if physicians react to financial incentives, say, by substituting highly reimbursed services for low-cost care, they no longer act as perfect agents. By studying how physicians react to a large reshuffling of their financial incentives, we add empirical evidence to this important literature. Our results may be informative to other settings with imperfect agents.

Finally, our results are of direct interest to policy makers. Reimbursement systems are periodically updated and should incorporate research insights to optimally construct fee structures that contain costs while maintaining quality.

The remainder of this paper is structured as follows. Section 3.2 discusses the institutional background and the payment reform of 2018. Our dataset is described in Section 3.3, and Section 3.4 provides descriptives. The empirical strategy is explained in Section 3.5, Section 3.6 presents and discusses the results. Section 3.7 concludes.

3.2 Institutional Background

The Swiss healthcare system is characterized by mandatory healthcare insurance (MHI) that covers a comprehensive and standardized basket of medical services. MHI accounts for the bulk of healthcare expenditures in the ambulatory sector. Insurance is organized according to the principles of managed competition (Enthoven, 1978). Individuals may choose among a number of insurance plans offered by private insurance providers. The basic plan for adults features a deductible of CHF 300 (≈USD 300) and free physician choice. In return for a premium reduction, individuals may opt for a higher deductible or the inclusion of managed care features.

Health plans may be changed during an annual open enrolment window in November.

Ambulatory care has traditionally been provided by independent single practices. Providers are reimbursed on a fee-for-service (FFS) basis (see Subsection 3.2.1). Most physicians are self-employed or work in small practices. In recent years, more group practices have emerged along with an increased importance of ambulatory clinics in hospitals that offer ambulatory services. Still, in 2014, 84% of ambulatory practitioners (co-)owned their current practice (Hostettler and Kraft, 2018). Despite the fact that most individuals enjoy free provider choice, the majority report having a regular GP (De Pietro et al., 2015).

3.2.1 Provider Reimbursement

Ambulatory services covered by mandatory health insurance are reimbursed under a universal fee-for-service (FFS) schedule.[4] Since 2004, fees are regulated on a national level in the *TARMED* fee schedule. The schedule contract is subject to collective negotiations involving the so-called *tariff parties*[5]. Fees for the roughly 4,500 services vary by physician specialty and region.

Similar to Medicare, medical services are ranked on a relative value scale (*Taxpunkte*) and conversed to a monetary value by multiplying with a geographic adjustment factor (*Taxpunktwert*).[6] Specifically, for service s, supplied in region r, a provider with specialty i may charge the following fee:

$$\text{fee(CHF)}_{s,i,r} = \text{relative value units}_{s,i} \times \text{geographic adjustment factor}_r \tag{3.2.1}$$

The relative value units of a service compensate the resources required for providing the service. First, medical work with the patient (*Medizinische*

[4]Note that exceptions prevail for patients insured with HMO plans. In these cases, insurance companies negotiate contracts with provider networks to arrange reimbursement of medical services. These contracts often include bonus payments or capitation mechanisms. Some physician networks may accept (partial) financial responsibility for the full clinical pathway of their patients (including inpatient care). Schmid et al. (2018) and De Pietro et al. (2015) offer further information. We abstract from individual contracts in this paper.

[5]Collective negotiations take place between the payer and the provider side. The supply side is represented by the Swiss Medical Association (FMH) and the Swiss Association of Hospitals (H+). The demand side is represented by two organizations representing the MHI insurance companies (santésuisse and curafutura).

[6]This factor varies by canton. Switzerland is a federal state consisting of 26 cantons. In 2017, these factors ranged from 0.82 (Schwyz, Zug) to 0.97 (Jura).

Leistung), and, second, the infrastructure, practice expenses, and overhead costs (*Technische Leistung*). The two components are explained in more detail in Appendix 3.A.

3.2.2 2018 Reform

Our analysis uses a large-scale change to the fee schedule that changed relative prices for physicians in 2018. Usually, the so-called *tariff parties* negotiate this national schedule. In 2018, the Federal Council intervened because the regulations for the current schedule expired and the involved parties had failed to negotiate a new contract. The reform affects the remuneration of services by changing fees, changing the composition of certain services, or retiring services.[7] In the end, fees were lowered for most services; the reform decreased the average fee by 14% (from CHF 400). Importantly, fees were also increased for some 30% of services[8] (accounting for 54% of total revenue in 2017), most notably consultation-related fees.[9] Figure 3.1 illustrates the fee changes.[10] From a physician's perspective, the revision changed relative prices and, depending on the specialty, affected a substantial part of practice revenue.[11] The new fees were announced only three months in advance, thus taking many physicians by surprise. Consequently, the 2018 Tarmed revision constitutes a natural experiment of reimbursement in a fee-for-service system that allows an in-depth analysis of how physicians respond to changes in financial incentives.

[7]Overall, services making up for a share of 27% of total revenue are affected by *compositional changes*. These changes entail five broad categories: First, splitting services into a finer set of services that are very similar to the old ones. These services are included in our models as they are tractable over time. Second, similarly, services were split into new services which are not comparable to the old service. Third, changes in the billable time of services with a *Minutage* (i.e. 5-minute intervals replacing 15-minute intervals). Fourth, introduction of *Minutage* for services. Fifth, changing the restrictions on the combination and repeated charges for a single service within a period of time (i.e. a maximum of four times per day). See Appendix 3.B for more details.

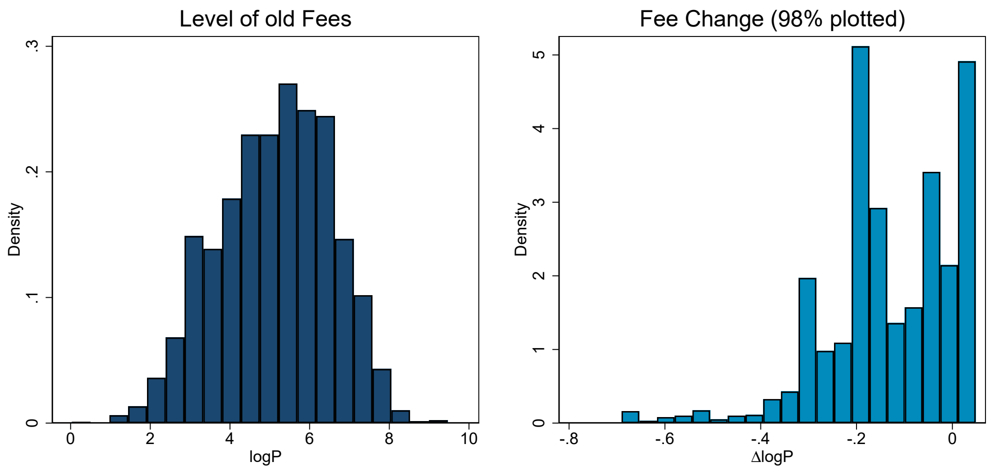
[8]Fee increases are mostly due to a harmonization of the scaling factor that compensates for the duration of medical training.

[9]Fee IDs 00.0010, 00.0020 and 00.0030.

[10]Note that the figure does not account for the importance of a service in terms of, say, revenue. Rather, all services are weighted equally to get a comprehensive picture of the fee changes.

[11]As a news article states (swissinfo, 2018): "At the end of January [2018], surgeons in Geneva went on a partial strike to protest against the lowering of tariffs (TARMED) for outpatient fees, managed by the Federal Office of Public Health. The surgeons refused to perform certain surgeries like carpel tunnel hand operations. This kind of action was unprecedented in Switzerland".

Figure 3.1: Distribution of Pre-Reform Fees and Fee Changes



Note: Both figures are based on the estimation sample. To increase readability, the right panel excludes outliers (bottom and top percentile).

Overall, services making up for a share of 27% of total revenue are affected by compositional changes. These changes entail five broad categories[12], which are described in further detail in Appendix 3.B.

3.2.3 Exogeneity of the Fee Changes

We argue that the fee change was exogenously determined. There are several justifications to this. First of all, the fee changes do not correlate with the lagged changes in service volumes (see Appendix Figure 3.6). Thus, the federal government did not systematically reduce the fees for services that grew in size. Several technicalities of the revision process further support this claim. Physicians were not directly involved in the process determining the new fees. Thus, they had no opportunity to guard important services strategically from fee cuts. Moreover, the revision consists of several different interventions such that physicians could not perfectly anticipate

[12]First, splitting services into a finer set of services that are very similar to the old ones. These services are included in our models as they are tractable over time. Second, similarly, services were split into new services which are not comparable to the old service. Third, changes in the billable time of services with a *Minutage* (i.e. 5-minute intervals replacing 15-minute intervals). Fourth, introduction of *Minutage* for services. Fifth, changing the restrictions on the combination and repeated charges for a single service within a period of time (i.e. a maximum of four times per day).

how their practice would be affected. Finally, the original *TARMED* fee schedule is based on pricing models that are intellectual property of the physicians' association. The federal government was not granted access to them.[13] Therefore, the new fees are not based on the same observables that determined the previous fees in the first place. We conclude that the fee changes are plausibly exogenous from the perspective of individual physicians.

3.3 Data

We use billing data from MHI in Switzerland for the period from 2015 to 2018.[14] The data, based on individual insurance claims, is aggregated on a physician-and-service level. It covers all physicians working in the outpatient sector who bill for MHI, excluding staffed physicians in hospitals.[15] We further aggregate all information on a quarterly frequency to achieve a physician-service level panel dataset. Using quarterly intervals is attractive because it reduces both noise and the computational burden. Further, there are fewer seasonal patterns. The dataset is best described by starting with the main variables: the quantities and the fees. Volumes are measured by how often a physician billed a service in a given quarter. Additionally, we observe how many patients a physician sees per quarter. The data distinguishes 42 medical specialties that were diversely affected by the fee changes. In the analysis, we will separately estimate the effects for the 12 largest specialties in terms of revenue[16] to account for potential heterogeneity. Taken together, they account for more than 80% of total revenue. An attractive characteristic of the data at hand is that it covers patients of virtually all health insurance companies in Switzerland.

[13] FOPH (2017, p. 10)

[14] We additionally have access to data for 2014. We exclude these data points from the analysis because of an earlier revision in October 2014. This reform aimed at redistributing revenue from specialists to GPs. We thus start our sample period in the last quarter of 2015. The quantities of 2014Q1 to 2015Q3 are only used to predict future quantities for the revenue decomposition model in Section 3.5.

[15] Outpatient services provided by hospital staff account for roughly 35% of total ambulatory costs under MHI. We exclude the hospital sector because information is not available on an individual physician level. Rather, the information in our data reflects an aggregate measure of decisions by several staffed doctors. Including them would thus conflate our estimates of individual behavior.

[16] In terms of revenue, the 12 largest specialty groups are GPs, psychiatrists, group practices, ophthalmologists, OB-GYNs, radiologists, pediatricians, medical practitioners, cardiologists, dermatologists, gastroenterologists and otorhinolaryngologists (ordered by revenue).

In order to increase the reliability of our analysis, we limit our sample to continuously practicing physicians[17] (excluding 17.5% of total revenue from 2015 to 2018) and exclude potentially erroneous data points. This includes services charged at highly unreasonable prices[18] (0.5%). Further, we restrict our analysis to services that can clearly be tracked across the different versions of the fee schedule. As discussed in Section 3.2 and Appendix 3.B, a number of services underwent major changes that render before-and-after comparisons unfeasible. We exclude these *compositional changes* (10.8%)[19] and retired services (0.2%). We will refer to the remaining data as estimation sample.

3.4 Descriptives and Graphical Evidence

This section provides descriptive statistics for the estimation sample used in the ensuing causal analysis. These are complemented by first graphical evidence of the reform in Subsection 3.4.1.

Table 3.1 provides summary statistics and some additional characteristics of our estimation sample from 2015 to 2018. In 2017, the last year before the reform, the sample consists of 13,051 physicians and 2,105 different services. It is important to point out that individual service composition is much narrower in scope: the average physician charges only 36 distinct services in a quarter. Notably, this number varies considerably across specialties, ranging from as few as six different services for psychiatrists to as many as 82 for radiologists. The average physician in our data charges a volume of 4,667 services and gains a revenue of CHF 90,239 per quarter. On average, a physician has 769 consultations per quarter.

A first descriptive analysis confirms the observation that costs rise over the years, which is consistent with an underlying trend of increasing healthcare expenditures OECD (2019). Quarterly revenue per physician increases from CHF 85,656 in 2015 to 90,239 in 2017 (see Table 3.1). In aggregate, this corresponds to a cost increase of 14% (from CHF Mio. 1,031

[17]We exclude physician-quarters if the revenue of the current or the preceding quarter lies below CHF 20,000 (25% of average quarterly revenue). Further, we excluded all observations of a physician if she had four or more quarters with total revenue below CHF 20,000. Finally, all observations of a physician were left out if she experienced extreme changes of her revenue in two or more quarters.

[18]In practice, we exclude physicians whose average prices deviate more than 10% from the official fees, or who deviate more than 50% from the fee at least once.

[19]Note that this value is lower than in Section 3.2 as the largest category of services affected by compositional changes is tractable over time and is thus still included in our models.

per quarter in 2015). The empirical analysis will have to account for this positive trend. We discuss the cost evolution in more detail in Subsection 3.4.1.

Table 3.1: Descriptive Statistics, Quarterly Means

	2015	2016	2017	2018
Number of Quarters	1	4	4	4
Physician Statistics				
Revenue (CHF)	85,656	85.725	90,239	91,919
	(135,588)	(139,849)	(150,878)	(152,306)
Quantity	4,492	4,484	4,667	4,784
	(5,769)	(5,854)	(6,203)	(6,916)
Consultations	708	792	769	837
	(722)	(790)	(779)	(919)
BITE				0.010
				(0.044)
Distinct Services	36	36	36	36
Aggregate Statistics				
Cost (Million CHF)	1,031	1,074	1,178	1,198
Predicted Cost (Million CHF)				1,262
Number of Physicians	12,031	12,528	13,051	13,036
Distinct Services (overall)	2,112	2,109	2,105	2,170
Observations	432,739	1,784,300	1,862,569	1,879,592

Notes: Panel A of this table reports quarterly means for physicians in our estimation sample. Panel B reports aggregate statistics for the estimation sample. Standard deviations in parentheses.

3.4.1 Graphical Evidence

This subsection provides first descriptive evidence of the effects of the policy change. For this purpose, we momentarily extend the scope of our analysis to the entire ambulatory sector covered by MHI, therefore going beyond our estimation sample described above. Specifically, the following discussion includes hospital staff and physicians who may have stopped or just started practicing over the course of our observation period. We describe how aggregate costs evolve and compare them to the announced policy target of saving 0.47 billion.[20]

We start by discussing how costs in the ambulatory sector evolve. Figure 3.2 shows total outpatient costs reimbursed by MHI for the years 2014 to 2019 (solid red line). Mirroring the previous discussion of Table 3.1, Figure 3.2 shows that overall medical costs grow over time. Ignoring this positive time trend would bias the estimated effects of the policy intervention.

Additionally, it is evident from Figure 3.2 that the data is cyclical: Ambulatory costs usually peak in the first quarter, followed by a through in the third quarter. This variation reflects an underlying seasonality in the prevalence of diseases such as the flu, which is amplified by missing visits during summer vacations.[21] We will account for such cycles by including physician-quarter fixed effects.

In the next step, we get a first notion of the policy's effect. To account for the high degree of cyclicality and the trend present in the data, we first build a simple counterfactual. We predict physician-specific service volumes based on pre-reform data under the assumption that quantities grow at the same rate for all physicians in the same specialty.[22] Multiplying by the pre-reform fees and aggregating over all physicians yields a notion of how medical costs would have evolved without the reform (dashed blue line). Comparing realized costs (solid red) to our counterfactual suggests that medical costs dropped considerably after the reform.

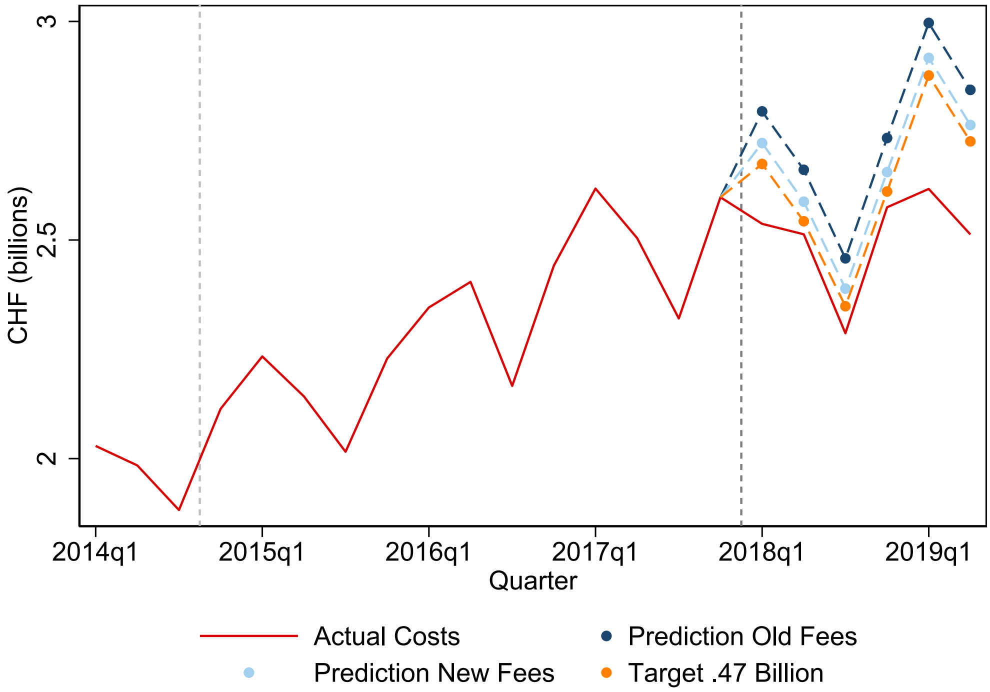
We suggest that part of this cost reduction is due to adjustments in physician behavior after the reform. Of course, part of the cost reduction is purely mechanical. After all, the reform reduced the average fee. In fact, the mechanical reduction only accounts for a relatively small share of the reduction: Multiplying counterfactual volumes by the *new* fees gives an indication of how costs evolved had physician behavior remained at predicted levels (subject to time trends and cyclicality). These costs (dashed pale blue) lie above the realized cost, suggesting that the cost reduction

[20]For the sake of completeness, Figure 3.8 in Appendix 3.F shows the same figure based on our estimation sample only. The general idea is comparable.

[21]A vast literature in public health and epidemiology shows increased mortality in winter months (e.g. Anderson and Bell, 2009; Wilkinson et al., 2004). It is natural to assume that this excess mortality coincides with increased utilization of medical services in the same period. Further, a literature in epidemiology shows that the incidence of several (infectious) diseases is highly seasonal. For example, in the northern hemisphere the seasonal flu occurs in the winter time (Dowell and Ho, 2004) and Chickenpox in spring (London and Yorke, 1973), thus a large share of both disease falling to the first quarter. Further, cold weather increases respiratory and heart problems, leading to increased utilization of primary care (Moineddin et al., 2008), and icy conditions increase the number of falls and accidents.

[22]Details on the predictions can be found in Appendix 3.D.

Figure 3.2: Total and Predicted Costs (Quarterly)



Note: The figure shows aggregate quarterly revenue for the outpatient sector covered by MHI. The dashed vertical lines mark the *TARMED* reforms. The figure is based on cleaned data including irregularly working physicians, hospitals and services affected by compositional changes.

is not purely mechanical but rather an additional reduction in costs due to behavioral changes.[23] This finding is somewhat puzzling. For instance, our interpretations runs contrary to the Medicare prediction of volume extension after fee cuts.

Did the policy attain the target of saving 0.47 billion? A back-of-the-envelope calculation suggests so. Actual costs lie below those predicted by a counterfactual that exactly achieves the saving target (dashed orange line), thus saving at least the announced target sum. To sum up, Figure 3.2 suggests that physicians responded to the reform beyond the purely mechanical effect. We will analyze their behavior empirically in the following sections.

[23]Interestingly, the discrepancy is most substantial in the first quarter of 2018, which could indicate that physicians need some time to adapt to the new fees optimally.

3.5 Empirical Strategy

The empirical analysis aims to identify how physicians react to the exogenous fee changes presented in Section 3.2.2. As mentioned earlier, there are at least two potential channels for physicians' response to fee changes: The so-called substitution effect and the income effect. The former is a volume increase (decrease) for services that became relatively more (less) attractive, and the latter is a volume increase of *all* services to achieve a targeted income. The remainder of this section explains how we isolate and analyze the two channels.

Measuring the bite of the reform. To analyze either channel, it is crucial to first measure how physicians are affected by the revision. Similar to Yip (1998)[24], we define a reform *BITE* variable for each physician - measuring the expected revenue loss related to the reform[25]:

$$\text{BITE}_{i,t} = \frac{\sum_s^S (P_{i,s,18} Q_{i,s,17t} - P_{i,s,17} Q_{i,s,17t})}{\sum_s^S (P_{i,s,17} Q_{i,s,17t})}. \tag{3.5.1}$$

For physician i in quarter t:, $Q_{i,s,y,t}$ represents the volume of service s in year y at fee $P_{i,s,y}$. *BITE* measures the expected percentage change in practice revenue based on pre-reform volume for each physician and quarter. A negative value thus indicates an (expected) income loss. In our view, our measure improves on earlier work by precisely estimating how revenue would change if quantities remained fixed. According to theoretical considerations in McGuire and Pauly (1991), physicians targeting a level of income would expand their total supply when they lose revenue in order to maintain their target income. In our setting, this prediction translates to a volume expansion when $BITE$ gets increasingly negative.

BITE ranges from -0.47 to 0.35, which highlights how widely the reform's impact on revenue varies between physicians. Absent any adaptions in behavior, physicians can expect a reduction of revenue by 47% while others would expect a 35%-gain in revenue. For the largest 12 specialties (in terms of revenue), the average (0.014) lies below the median (0.034),

[24]Yip (1998) proposes a weighted average of the fee changes. The weights are computed as the average quantities right before and after the treatment. In fact, our measure in (3.5.1) highly correlates with an alternative specification in the style of Yip (1998) (correlation coefficient 0.71).

[25]Note that we want to build a comprehensive measure of how physicians are affected. Therefore, we also include services affected by compositional changes to build this variable.

which implies that there are fewer losers than winners, and that losers are more strongly affected in absolute terms. Tables 3.2 and 3.3 in Appendix 3.E show summary statistics for the variable by medical specialty.

3.5.1 Substitution Across Services

How does the (relative) provision of different services respond to changes in their relative prices? We identify the substitution pattern by comparing relative changes in service quantities *within* a physician and quarter. Formally, this relative change is modeled by including a physician-quarter fixed effect in the regression model. The underlying rationale is that fee changes for the same service may have different consequences for physicians who differ in their practice style, even if they practice in the same specialty. In turn, the analysis abstracts from explicitly modeling the effect of fee changes on a physician's overall service volume.[26] Our baseline specification of the model is as follows:

$$\Delta \ln Q_{i,s,t} = \beta_0 + \beta_1 \Delta \ln P_{s,y} + \gamma_s + \delta_{i,t} + \epsilon_{i,s,t}, \tag{3.5.2}$$

where $Q_{i,s,t}$ is the volume of service s provided by physician i in quarter t, $P_{s,y}$ represents the corresponding fee in year y, and Δ denotes the difference to the prior-year quarter.[27] Using the first difference of the logarithms of the outcome as well as our main regressor is attractive for two main reasons. First, the log-transformation facilitates comparison despite the large variance across services and physicians. Second, comparing to the prior-year quarter nets out quarter effects. We further include two fixed effects in our baseline specification: A service fixed effect (γ_s) that takes care of service-specific trends, and, as already motivated above, a physician-quarter fixed effect ($\delta_{i,t}$) that allows for comparing relative changes within a physician and quarter. Note that the physician-quarter fixed effect ($\delta_{i,t}$) picks up the *BITE* variable and thus helps to isolate the substitution effect. Our coefficient of interest, β_1, will (for small changes) approximate the percentage point change in service volume when the relative fee changes by one percentage point.[28]

[26] In fact, our preferred specification does not allow for the concurrent modeling of both, as average changes are picked up in the physician-by-quarter fixed effect.

[27] Note that the fee schedule mandates an upper limit for the fees. Technically, physicians are allowed to charge lower fees, which introduces individual-specific variation. In practice, we take care of this by using the average fee charged by a physician before and after the reform. We omit the i subscript for the fees for the sake of readability.

[28] To be precise, $\beta_1$measures how a change in the log gross growth rate of $P_{i,s,y}$ affects the log gross growth rate of $Q_{i,s,t}$. For small changes of $Q_{i,s,t}$ and $P_{i,s,y}$, the coefficient

Our estimate of β exploits variation across time and between services. In order to claim that β identifies the causal effect of $\Delta\ln P_{s,y}$ on $\Delta\ln Q_{i,s,t}$ in Equation (3.5.2), we need to establish that $\Delta\ln P_{s,y}$ is unrelated to $\epsilon_{i,s,t}$. More precisely, we need a conditional independence assumption (CIA) of the following form:

$$E[\epsilon_{i,s,t}|\Delta\ln P_{s,y}, \gamma_s, \delta_{i,t}] = E[\epsilon_{i,s,t}|\gamma_s, \delta_{i,t}]$$

Under the CIA, no relation remains between $\Delta\ln P_{s,y}$ and $\epsilon_{i,s,t}$ once accounted for service-specific trends and physician-quarter fixed effects. Because the model is estimated in first differences we additionally account for observed and unobserved time-constant service characteristics (such as treatment duration, technological requirements) and physician characteristics (such as a physician's practice style, her pool of patients, or local healthcare market conditions). While the CIA is fundamentally untestable, we argue that it is likely to hold. As argued in Subsection 3.2.3, the nature of the fee changes was difficult to anticipate for the physicians and they were not directly involved in the negotiations. Taken together, we account for a great deal of dynamics by first differencing and introducing a linear service trend (γ_s) as well as quarterly fixed effects on the physician level ($\delta_{i,t}$). Thus, it is unlikely that the fee changes are linked to unobserved time-varying characteristics in equation (3.5.2). Note that the identification of the causal effect of $\Delta\ln P_{s,y}$ is based on variation over time and variation across services. This renders the identification strategy robust against contemporaneous events as it is rather the physician's distribution of services that identifies the treatment effect.

The baseline model in Equation (3.5.2) is appealing for several reasons. First, it represents the only way to isolate the substitution effect neatly.[29] To the best of our knowledge, our paper is the first to draw conclusions about the substitution pattern shutting other response channels. Further, another attractive property is that the coefficient of interest, β_1, has the flavor of an

approximates net growth rates. Recall that the gross growth rate of $Q_{i,s,t}$ with respect to the prior-year quarter is $\frac{Q_{i,s,t}}{Q_{i,s,t-4}}$ whereas the net growth rate subtracts 1 from the gross growth rate.

[29] The physician-quarter effect makes it a model of relative fee changes for each physician, which we ultimately want to estimate. In principle, the effect of $\Delta\ln P_{i,s,t}$ and $BITE_i$ could be estimated in one regression by dropping the physician-quarter fixed effect. However, the same value of $\Delta\ln P_{i,s,t}$ could have very different implications, say, because for a physician all other fees may have increased or decreased.

elasticity.[30] Finally, the estimation model presented in Equation (3.5.2) ties to the previous literature (e.g. Yip, 1998), thus ensuring comparisons. To the best of our knowledge, no previous paper has modelled substitution effects across such a broad range of services, and for such a heterogeneous group of physicians. We thus contribute novel evidence to the literature.

To ensure that we capture the appropriate effect of the fee variable ($\Delta \ln P_{i,s,t}$), we additionally consider more flexible specifications such as introducing polynomials of the fee (up to order five) and interacting with an indicator for a fee decrease. We use the physician-and-service-specific revenue of the prior-year quarter as weights for all regression models. More precisely, we use the logarithm of the revenue to be consistent with our main outcome and treatment variables. All standard errors are clustered on the physician-quarter level.

As discussed in the first part of this section and illustrated by Tables 3.2 and 3.3 in Appendix 3.E, the bite of the fee reform varies considerably across medical specialties. Thus, in the next step, we consider if, in turn, heterogeneous substitution patterns arise. Formally, we let β_1 in Equation (3.5.2) vary for the 12 largest specialties and group the results by the specialty's median *BITE*.[31] Generally speaking, the incentives and possibilities to substitute may depend on a number of factors. Most importantly, we expect physicians who have lost a large share of their revenue to respond more strongly for two reasons. First, markedly lower fees imply a decrease in revenue. Because this increases the marginal utility of revenue under a range of utility functions,[32] physicians have ample incentives to expand the volume of services with a higher financial return. Such behavior manifests in substitution from "losing" to "winning" services in terms of the relative fee. Second, experiencing a severe reduction in revenue is correlated with a higher variance of the fee changes[33], which means that relative prices, and

30

$$\beta_1 = \frac{\partial \Delta lnQ}{\partial \Delta lnP} = \frac{\partial ln\left(\frac{Q_t}{Q_{t-4}}\right)}{\partial ln\left(\frac{P_t}{P_{t-4}}\right)} = \frac{\frac{Q_t}{Q_{t-4}}}{\frac{P_t}{P_{t-4}}} \cdot \frac{\partial \frac{Q_t}{Q_{t-4}}}{\partial \frac{P_t}{P_{t-4}}}$$

31 Taken together, these physicians account for more than 83% of all revenue. The included specialties are GPs, psychiatrists, group practices, ophthalmologists, OB-GYNs, radiologists, pediatricians, medical practitioners, cardiologists, dermatologists, gastroenterologists, and otorhinolaryngologists (ordered by revenue).

32 This claim holds for standard utility functions as outlined in McGuire and Pauly (1991).

33 Figure 3.9 in Appendix 3.F provides evidence for this claim.

thus incentives, have changed more strongly.

Besides the possibly amplifying effect of the income $BITE$, there are constraints on how physicians can optimize their service mix. An apparent one is the number of distinct services a physician uses: Having fewer distinct services limits the substitution possibilities. Further, the fee schedule limits how often certain services can be charged per patient and consultation. It seems likely that expanding service quantities is harder when such limitations restrict a larger part of a physician's practice.[34] We use the revenue-share of limited services as a rough proxy for the ease of substituting. Because the variable is constant for each physician and quarter, the physician–quarter fixed effect ($\delta_{i,t}$) in Equation (3.5.2) picks it up.

3.5.2 Total Healthcare Supply

In a second step, we focus on a second response channel: volume expansion. Several earlier studies (e.g. McGuire and Pauly, 1991; Yip, 1998) found that physicians with the sharpest income decline show the most pronounced reactions. Thus in a second analysis, we examine how a physician's *total* healthcare supply responds to income changes. This is a non-trivial exercise for two reasons. First, the raw quantities of the services are not directly comparable across services and physicians. Thus, they cannot simply be summed up to get a measure of total supply. Second, services exhibit different time trends, which renders the analysis of any aggregate delicate, as it remains unclear how these trends should be averaged. We tackle these issues by proposing two complementary identification strategies.

In a first step, we argue that the number of office visits constitute an appropriate measure of a physician's healthcare supply. Thus, we regress the change in consultations in a given quarter on the individual, quarterly $BITE_{i,t}$, and a physician fixed effect. We let the coefficient vary for medical specialties with a median revenue loss (negative BITE), indexed by $m \in M$.

$$\Delta \ln C_{i,t} = \beta_0 + \beta_1 BITE_{i,t} + \sum_{m \in M} \beta_{1,m} I_m \times BITE_{i,t} + \delta_i + \epsilon_{i,t}. \qquad (3.5.3)$$

In contrast to the model of the previous section, the identification strategy of this approach purely relies on variation over time. Essentially, we assume that, once accounted for the average and the trend of a physician's number of consultations, $BITE_{i,t}$ is unrelated to any time-varying unobservables, i.e. $E[\epsilon_{i,t}|BITE_{i,t}, \delta_i] = E[\epsilon_{i,t}|\delta_i]$. The assumption seems plausible because,

[34] Figure 3.10 in Appendix 3.F summarizes the number of services, the fraction of revenue restricted by limitations, and the median value of BITE for the largest specialties.

again, the fee changes were difficult to anticipate from a physician's perspective. Thus, a correlation of $BITE_{i,t}$ with time-varying characteristics seems unlikely. For most specialties this claim is supported by the means of Figure 3.11 in Appendix 3.F, as pre-reform consultation growth does not seem to depend on $BITE_{i,t}$. Unfortunately, for psychiatrists, gynaecologists, and medical practitioners, pre-reform growth rates of consultations seem to correlate with the value of $BITE_{i,t}$. Thus, the model in (3.5.3) encounters an endogeneity problem for these medical specialties. This undermines the identification of β_1 in (3.5.3) and makes the estimation of the model for these specialties infeasible. We thus refrain from estimating the model for these three specialities.

Further, it seems implausible that an unobserved but contemporaneous event (heterogeneously) affected the number of consultations, and was at the same time correlated with $BITE_{i,t}$. For instance, consider an unexpected shift in the consumer's demand for medical services from 2017 to 2018. It would need to be the case that this shift is correlated with $BITE_{i,t}$ which is unlikely to hold as consumers are weakly affected by the reform due to their limited payment obligations.

Identifying β_1 and $\beta_{1,m}$ in equation (3.5.3) relies on sufficient variation of $BITE_{i,t}$ within the specialties.[35] Table 3.2 suggests that physicians may be very heterogeneously affected by the reform even within specialties. Thus, we argue that there is enough variation to identify β_1 and $\beta_{1,m}$.

We expect only the physicians affected by a substantial revenue reduction to exhibit income-targeting behavior. Besides, consultations have been growing at different rates for these specialties. Thus, it is important to estimate a fully saturated model. Comparing to the substitution model in (3.5.2), it is apparent that the model in Equation (3.5.3) builds on fewer observations because the units of observation are providers instead of providers-and-services.

The second approach is motivated by a prediction of the PID hypothesis (McGuire and Pauly, 1991). The PID states that physicians fully offset any income loss by increasing service volume. For this purpose, we will decompose the observed revenue change from 2017 to 2018 into two parts: First, a *mechanical change* that measures by how much revenue changes due to the fee adjustments, holding the supply level fixed. Second, a *behavioral change* measuring how revenue changes because the provision of services adapts (holding prices fixed). Note that individual physicians can not influence the first part. They may react, however, by, say, performing

[35]Technically, this is only true for the specialties in M. We identify β_1 through the pool of all physicians not belonging to M which further adds exploitable variation.

additional examinations, which the second part would pick up. Relating the two measures provides an estimate as to how much physicians offset lost income. Thus, the PID prediction (coefficient of -1) will be testable. In practice, we will base the measures on cost predictions that account for specialty-specific trends and cyclicality. More precisely, we use specialty-specific quantity predictions $\hat{Q}_{i,s,18}$ for 2018 to account for different trends in volumes.

$$\Delta \ln R_i = \underbrace{\ln\left(\sum_s^S P_{i,s,18} \cdot Q_{i,s,18}\right) - \ln\left(\sum_s^S P_{i,s,18} \cdot \hat{Q}_{i,s,18}\right)}_{\text{behavioral change}_i} + \underbrace{\ln\left(\sum_s^S P_{i,s,18} \cdot \hat{Q}_{i,s,18}\right) - \ln\left(\sum_s^S P_{i,s,17} \cdot \hat{Q}_{i,s,18}\right)}_{\text{mechanical change}_i}. \tag{3.5.4}$$

$\ln\Delta R_i$ measures how the revenue in 2018 compares to what was expected for a physician based on previous quantities (levels and trends) and the old fees. The first term of the decomposition identifies how much of $\ln\Delta R_i$ is due to a change in service volume (*behavioral change*). Similar to the $BITE$ variable, the second term (*mechanical change*) captures how revenue (mechanically) changes due to changing prices. Compared to the $BITE$, the *mechanical change* additionally accounts for specialty-specific trends in service volumes. Again, the variable further improves on the original definition of $BITE$ by Yip (1998) by using quantity predictions that are unaffected by physicians' responses to the reform. The predictions for the quantities in 2018 are based on the assumption that service-specific growth rates are constant across all physicians of the same specialty. Further, owing to the data's high cyclicality, we use average quarterly differences[36] and not linear trends for the predictions. Appendix 3.D describes the predictions in further detail. Applying these counterfactuals allows linking revenue changes directly to (mechanical) changes in fees and (behavioral) changes in quantities. If physicians target an income level, they would choose their volumes in a way that fully offsets the revenue loss due to the fee changes and, in turn, effectively reduce the change in revenue to zero. Thus, regressing *behavioral change* on *mechanical change* provides an answer to whether this is the case or not.

The decomposition in Equation (3.5.4) constitutes a sophisticated comparison to quantify the correlation between the two components of revenue change on the right-hand side of Equation 3.5.4. We account for the

[36] Average change in quantity for all first, second, third, and fourth quarters.

physician's average revenue through first differencing and for the trend in revenue by considering the predictions $\hat{Q}_{i,s,18}$. Note that, in contrast to the previous models, we do not identify a causal effect in this estimation. The model is only estimated for one time period. Therefore, we cannot account for unobserved time-varying characteristics on the physician level. This is, it is infeasible to estimate an analogue of $\delta_{i,t}$ or δ_i in equations (3.5.2) and (3.5.3). Thus, we regard the revenue decomposition results as a descriptive supplement to the previous models. Finally, because the analysis is conducted using only one observation per physician, we perform a median regression to ensure that the results are robust to outliers (Hao and Naiman, 2007).

Both models in this subsection aim at quantifying how physicians adapt their overall healthcare supply once their projected revenue is altered by the reform. Beyond accounting for physician-specific trends, we want to add further credibility by including covariates approximating the ease of substitution and the pool of patients.[37] Both may limit how flexibly physicians can adjust their overall supply. More precisely, we include the number of distinct services charged by a physician. The reasoning being that regularly prescribing a larger variety of services likely increases substitution possibilities. In turn, physicians charging many different service may rather substitute than increase their overall supply. Further, we account for how much of a physician's revenue is due to treating patients who are young (below 36), female, morbid (hospital stay in the previous year) and/or have a low deductible. Ultimately, the patient pool may also restrict the ease of labor supply expansions. Finally, a high physician density (per specialty and canton) is likely to increase a consumers' choices and therefore limits the market power of the physician. This would make it more difficult to extend labor supply. Note that in the case of the decomposition exercise, we additionally include specialty fixed effects for all physicians $m \in M$.

3.6 Results and Discussion

3.6.1 Substitution across Services

Appendix Table 3.4 shows our baseline estimates of the causal effect of relative fee changes on changes in relative service volume. The magnitude of the effect is sizeable, with a coefficient of 0.253. In other words, when

[37]Note that in equation (3.5.3) we explicitly for physician trends in consultations while in equation (3.5.4) this is implicitly addressed by considering the predictions for the quantities $\hat{Q}_{i,s,18}$

the fee of a service increases by 2% more than the average, its volume will increase by around 0.5%. Figure 3.3 illustrates the result. The figure shows the predicted change in service volume based on Equation (3.5.2) in brown. The results show that physicians expand (reduce) the quantity of services for which the fee has become relatively more (less) attractive. Here, losing or gaining attractiveness is measured relative to the weighted[38] average fee change for every physician since we include physician-quarter fixed effects.

We estimate two additional models that allow for potentially different substitution responses for services with above- and below-average fee changes. The results, pictured in orange (kink) and red (quadratic fit), suggest that physicians respond more strongly to above-average fee changes, implying that they rather increase financially favorable services than reduce unattractive ones. We conclude that it is sensible to estimate a flexible model and thus will only present results from the quadratic fit in the remainder of this section. The estimates are robust to a range of more flexible specifications discussed in Section 3.5.[39]

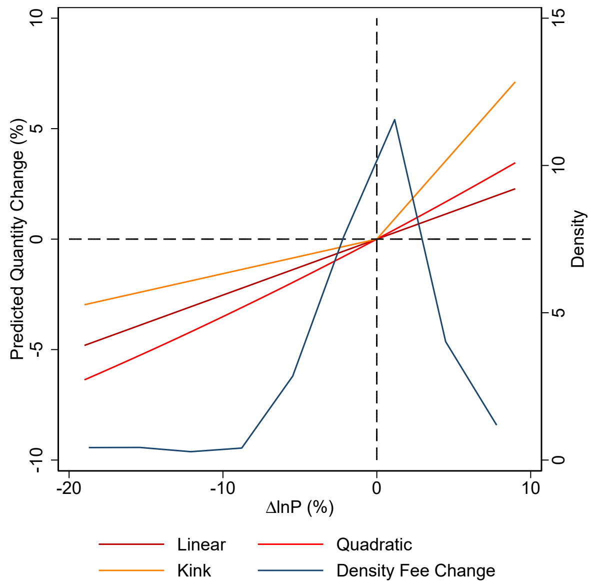
In terms of the direction of the effects, our results confirm earlier findings, but are slightly smaller in magnitude (Gruber et al., 1999). Comparisons need to be taken with a grain of salt since earlier studies estimated substitution between two different markets (Yip, 1998) or for only two different services (Gruber et al., 1999) and are thus not directly comparable. Nevertheless, the described substitution pattern is in line with theoretical models (McGuire and Pauly, 1991). The results present compelling evidence that physicians behave rationally and maximize their revenue. Potentially, physicians may maximize profits against the best interests of their patients and thus cause a welfare loss, for example by providing excessive care beyond the medically optimal level. However, the model does not allow a welfare statement, since the optimal level of care remains unobserved.

Heterogeneity. We next analyze how substitution patterns vary between different medical specialties whose revenue is differently affected by the reform. We thus interact the fee variables in (3.5.2) with a dummy for each of the 12 largest specialties. The results using the quadratic fit are illustrated in Figure 3.4 (and Appendix Table 3.5). To increase readability, the figure groups specialties by how strongly their revenue is affected. Specifically, we distinguish three groups based on the median *BITE*: *negative*

[38]The observations are weighted by their prior-year quarter log revenue.

[39]Estimation of specifications with higher-order polynomials yield qualitatively similar results. The remainder of the paper thus focusses on models including terms up to second-order polynomials.

Figure 3.3: Substitution Pattern



Note: The figure shows the predicted quantity change by fee change based on the estimated fee parameter(s) β_1 from three specifications: a linear term in the fee variable, complemented by an interaction of the fee with an indicator variable for a fee decrease (*kink*), or a quadratic term (*quadratic*). Observations are weighted with the prior-year quarter log revenue. The density represented in this graph is the pdf of the centered fee changes, weighted with the same weights used for the regression models.

bite (radiologists, cardiologists, gastroenterologists)[40], *no bite* (psychiatrists, group practices, otorhinolaryngologists, OB-GYNs, dermatologists), and *positive bite* (GPs, pediatricians, medical practitioners, ophthalmologists). The broad implication of the preceding model remains: physicians increase the quantity of services in cases where the fee has relatively increased. This is particularly true for providers who benefit from the reform (right part of the figure). However, the other two subfigures reveal that the specialties differ regarding the strength of the substitution response.

In the (largely) non-affected group (middle panel), the general picture still holds. However, psychiatrists show an opposite pattern, reflecting a seemingly irrational substitution pattern, with volume increases of relatively less attractive services. Possibly, this puzzling result arises due to constraints: psychiatrists only choose from six distinct services in the first place, and nearly all of them are limited (accounting for 98% of revenue).

[40]Bite takes on the lowest median values for radiologists (-0.130), cardiologists (-0.094), and gastroenterologists (-0.099).

Further, none of the relevant services charged by psychiatrists seem to be substitutes.[41] Being virtually unable to substitute, psychiatrists could potentially respond by expanding their aggregate supply of healthcare. Subsection 3.6.2 will tackle this issue.

Lastly, we turn to the group that loses revenue due to the reform (left subfigure). Cardiologists respond with a pronounced substitution behavior. This result is in line with our expectation that physicians react more strongly when they lose a larger share of their revenue. Further, Figure 3.4 illustrates that service volumes of radiologists and gastroenterologists are much less sensitive to fee changes, and therefore exhibit a less pronounced substitution response. The stark contrast of this result to the previous discussion raises the question if physicians who lose a large chunk of their income respond differently. It is possible that once a certain threshold is met, physicians care much less about relative prices but rather expand volumes across all services, irrespective of the fee changes. If this is the case, our results will not pick it up. Therefore, we relate volume and *BITE* directly in the following section.[42]

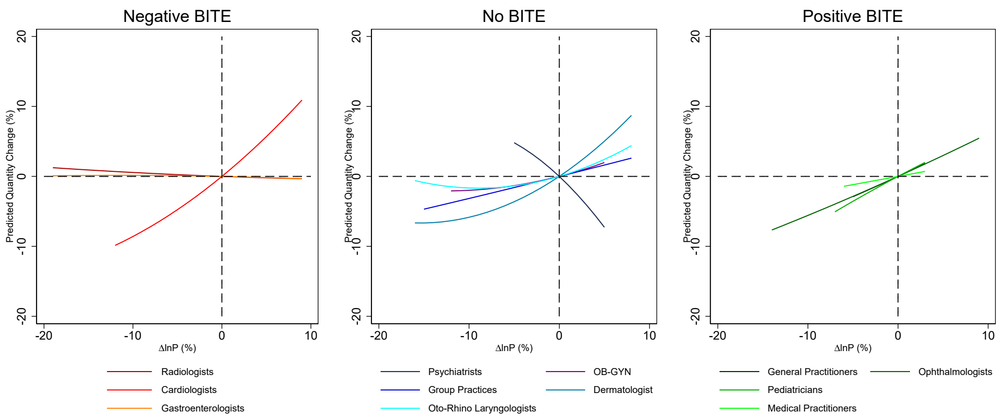
3.6.2 Total Healthcare Supply

The previous subsection established that physicians who suffer a financial loss substitute the least. Likely, the second channel is more important: an overall supply expansion. Among the heavily affected group, only cardiologists substitute weakly according to the expected pattern. However, radiologists, cardiologists, gastroenterologists and potentially psychiatrists may respond by increasing their *total* supply. Generally, volume can be

[41] Two services generate about 90% of psychiatrist revenue: 02.0020 and 02.0210, (both charging for therapy in five-minute intervals). While these positions are similar in scope, they differ in the practitioner. By law, MHI does not reimburse psychotherapists (without a medical degree) in independent practices. In order to be eligible for reimbursement, they need to be employed by a psychiatrist who files the claims on their behalf and supervises them. This regulation directly concerns the substitution possibilities for the two positions mentioned above. Specifically, 02.0210 can only be charged by employed psychotherapists (without a medical degree). Hence, the two positions are prescribed by two different medical professionals. This makes substitution across services difficult. A psychiatrist can only increase the volume of 02.0210 by shifting his workload to employees (or employing new psychotherapists). While this substitution is, of course, possible in reality, modeling substitution *across* suppliers lies outside the scope of this paper.

[42] Similar to psychiatrists, radiologists are restricted in their choices by facing numerous limitations (97% of their revenue). Cardiologists and gastroenterologists face almost no restrictions (26% and 13% of their revenue, respectively) and charge a moderate number of distinct services (25 and 34). Thus, these reasons may only partly explain the weak substitution response.

Figure 3.4: Substitution Pattern by Physician Specialty



Note: The figure shows the predicted quantity change by fee change for the twelve largest specialties. Predictions are based on the quadratic specification from Figure 3.3: the model includes a quadratic term for $\Delta \ln P$ that also varies for the twelve largest specialties. Observations are weighted with the prior-year quarter log revenue. Higher order terms do not alter the interpretation. The substitution pattern is shown for 90% of the fee changes within the physician specialty.

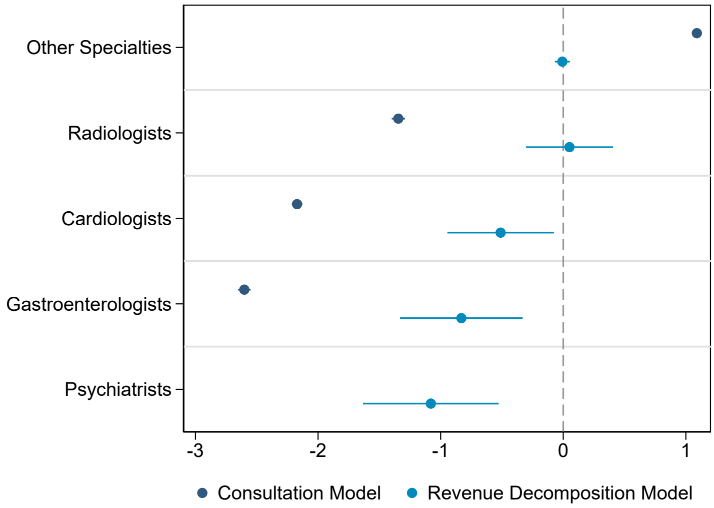
expanded along two margins: Increasing the number of consultations or charging more services per visit. Based on theoretical considerations in McGuire and Pauly (1991), we expect only physicians whose income is substantially reduced to increase their aggregate supply. Further, and in an extension to the theoretical model, we expect a volume increase for psychiatrists. As mentioned above, they constitute a group with very limited substitution possibilities so that their only margin is a supply change. In the following, we show and discuss estimation results for the aforementioned groups (negative median BITE and psychiatrists[43]) and compare them to a benchmark group consisting of all other medical specialties.[44]

Figure 3.5 presents point estimates and standard errors for the main coefficients from the two models laid out in Subsection 3.5.2: The coefficient of *BITE* in the consultation model in Equation (3.5.3) in dark blue, and the coefficient of *mechanical change* from the revenue decomposition model based on Equation (3.5.4) in light blue. A negative coefficient indicates that large revenue losses are (partially) offset by an increase in consultations or aggregate supply and thus provides evidence for income targeting. Note

[43] Note that for psychiatrists, the consultation model is not run due to the endogeneity concerns discussed in Subsection 3.5.2.

[44] As mentioned in Section 3.5, psychiatrists, medical practitioners, and OB-GYNs are excluded because of endogeneity concerns.

Figure 3.5: Total Healthcare Supply



Note: This Figure shows the slope coefficients of the models outlined in Subsection 3.5.2 (point estimates and 95% confidence intervals). Observations are weighted with either the prior-year quarter log revenue (consultation model) or the prior-year log revenue (revenue decomposition model).

that the standard errors are larger in the revenue decomposition model due to fewer observations (one observation per physician). The full set of results is shown in Appendix Tables 3.6 and 3.7.[45]

First of all, Figure 3.5 illustrates that physicians respond differently depending on whether they gained or lost due to the reform. We first discuss the benchmark group with an unaffected or even increased revenue. For physicians in this group, $BITE$ takes on a median value of about 0.03. Our results suggest that these physicians increase consultations in response to the (modest) revenue gain. The estimated coefficient corresponds to a 1.09% increase in consultations when the expected revenue (measured by $BITE$) increases by one percentage point. Considering the revenue decomposition exercise, the *mechanical change* has no effect on aggregate supply

[45]Robustness and sensitivity checks of those models include a set of regressions where we control for covariates. Specifically, we include quarterly physician data on the number of distinct services, the cantonal physician density, and the revenue share of services that face some limitations. Further, we account for differences in the patient groups by including the revenue shares of the following groups: women, young (below age 36), morbid (hospital stay in the previous year), low deductible (CHF 300). The qualitative implications for both models remain the same. However, for the revenue decomposition model, including covariates alters the statistical significance because the estimation model only includes one observation per physician.

for these physicians. These results speak against the literal target income hypothesis. Not a surprising finding, since the (literal) income-targeting hypothesis is generally deemed unlikely to hold for physicians who gain revenue (McGuire and Pauly, 1991).[46] Rather, the observed pattern may be explained by two competing explanations. On the one hand, the pattern is consistent with revenue maximization, as the increased fees increase the marginal return to labor, which would manifest in a positive coefficient of *BITE* in the consultation model. This is exactly what we observe for physicians in the benchmark category. On the other hand, the pattern may result from substitution. Because the reform increased consultation fees[47], office visits gained attractiveness relative to other services. Therefore what we see in Figure 3.5, i.e. seeing more patients but not expanding aggregate supply, may reflect a substitution behavior.

Next, we turn to the group with a negative *BITE*. Revenue losses induce cardiologists and gastroenterologists to expand both their number of patients and their total supply. The coefficients of the consultation model translate to an increase of up to 2.6% in the number of patients when *BITE* decreases by 1%-point. Aggregate supply measured by *behavioral change* is less responsive but increases by 0.51% and 0.83%, respectively, if *mechanical change* decreases by 1%. This result provides evidence for income-targeting behavior. In contrast, radiologists do not increase their service volumes. A possible reason for this difference is that a radiologist's practice mostly relies on referrals by other providers. Thus, radiologists have fewer options to increase their supply. Another explanation could be that they have one of the highest quarterly revenues among all physicians in which case a standard utility function implies that the marginal benefit of revenue is lower.[48] Thus, a comparable reduction in revenue does not lead to the same response as revenue matters less for overall utility. However, radiologists increase the number of patients they see, which likely corresponds to an increase in labor supply. Therefore, and similar to psychiatrists' limited scope for substitution, radiologists may be constrained with respect to the services they charge *per patient*.

Finally, psychiatrists expand the quantities of their services (coefficient

[46] In principle, physicians who benefit from the reform (positive *BITE*) could reduce their supply. This behavior is consistent with the literal target income hypothesis of McGuire and Pauly (1991). However, this behavior is generally regarded as highly unlikely even by the authors. Thus, we do not expect a negative coefficient for physicians with a positive $BITE$.

[47] Service IDs: 00.0010, 00.0015, 00.0020.

[48] The median quarterly revenue of radiologists is about CHF 380,000. Among the largest 12 specialties, the second highest value is CHF 225,000 for group practices.

of the revenue decomposition model: -1.08). As described in Subsection 3.6.1, psychiatrists exhibit a puzzling substitution pattern that favors less attractive services. Taken together, the results suggest that psychiatrists provide more healthcare and consequently partially offset their revenue loss. Again, this observation is consistent with the income-targeting hypothesis and thus with physician-induced demand.

3.6.3 Which Channel is More Important?

So far, we have found evidence for both a substitution response and an adjustment of overall healthcare supply.Which response is more important in terms of healthcare costs? We can't directly answer this question using the results described above. The underlying estimation models differ in the observation unit and outcome variable (see Section 3.5). To better compare the two effects, we employ a unified decomposition model, similar to Equation (3.5.4). In a nutshell, the model attributes changes in revenue to (i) fee changes, (ii) labor supply changes, and (iii) a modified service mix. Appendix 3.C introduces the extended decomposition model.

In line with our main findings, the results suggest that physicians increase expand volumes (via labor supply) in response to a substantial revenue loss. Physicians who suffer from a 5% (mechanical) revenue loss increase their supply by roughly 3%. Contrastingly, the substitution response does not translate to a substantial effect on revenue. After all, physicians may be constrained by their pool of patients, their chargeable set of services, or the newly introduced rules and limitations.[49] We conclude that, at least in the short run, the increase in total healthcare supply (volume expansion) is most relevant in terms of revenue, and thus, healthcare costs.

Taken together, we observe four results. First, in response to a large-scale fee change, physicians substitute across services to optimize their revenue. Second, physicians respond differently depending on whether they gained or lost revenue. Being hit harder by the reform, most physicians tend to increase their aggregate supply rather than to substitute between services. Third, the ability to substitute may depend on other factors such as the number of services a physician regularly charges, and whether these services are close-enough substitutes. Finally, in terms of overall costs, the healthcare supply channel matters most. This result confirms earlier

[49]Note that this result does not contradict the previous findings related to the substitution channel. As we included fixed effects into the regression model in (3.5.2), we are essentially comparing relative changes in quantities. Yet, the average changes may be substantial. Adding to this, it is a priori unclear how to conceptualize an aggregate measure of the substitution response.

findings by Burkhard et al. (2019) who find that the volume change is more important than substitution in explaining cost differences between physicians who sell drugs from a practice pharmacy and others who are only prescribing the medicine.

3.6.4 Limitations

Our results are subject to a number of limitations. Most importantly, the dataset does not allow conclusions regarding patient health. Consequently, we form no statement if and how patient health is affected by the changes in the service mix described in Subsection 3.6.1 or by the changes in treatment intensity described in Subsection 3.6.2. Further, as in any empirical study of physician behavior, the optimal level of care remains fundamentally unobserved.

Moreover, we need to exclude a subset of medical services due to data constraints. These services, mostly affected by compositional changes are not directly comparable to pre-reform services. Therefore, the substitution analysis in Subsection 3.6.1 does not include the full variety of available services. However, the excluded services only make up around 10.8% of total revenue, and we therefore do not expect a large impact on the estimated results. Excluding the services potentially biases our results if they are systematically differently substituted, which we do not expect. Similarly, we observe physician behavior within the scope of MHI. For the ambulatory sector, MHI accounts for the vast majority of health expenditures. Still, services not covered by MHI may be alternative substitution options that remain unobserved to us.[50]

3.6.5 Policy Implications

Our results provide compelling evidence that financial incentives affect treatment decisions. First, we find substitution across the board. Second, the response varies considerably by the extent of revenue at risk. These findings have important implications for the optimal design of remuneration systems. Health policy is frequently confronted with the task of finding better ways to contain ever-growing healthcare costs while maintaining quality. Most regulators mandate regular updates of the provider payment systems, which offer frequent opportunities to incorporate new evidence

[50]In multipayer markets, fee changes may also spill over to both prices and volumes of other payers (e.g. Clemens and Gottlieb, 2017, for the US). Our setting features a single payer so that no outside option persists.

on the provision of medical services. We argue that these periodic updates should also consider economic insights into the responses to financial incentives. Our findings suggest that aggregate supply increases if a sizeable part of revenue is lost. We thus conclude that gradual fee changes are preferable since strong and costly responses due to the income loss can be avoided.

Changes in relative fees may lead to potentially undesirable substitution between services. In principle, the regulator could shut the substitution channel. Reducing all fees by the same factor keeps relative prices unchanged. Careful considerations may additionally keep the supply extension at a minimum by keeping the income effect small. Of course, this is a rather technical argument, because in practice, many other factors influence the decision on fee adjustments. Perhaps a more realistic approach in this direction is trying to identify a medical specialty's most relevant set of services and keep their relative fees relatively constant.

Further, from a long-term perspective, many countries discuss more large-scale reforms of the payment *system*, for example, by abandoning FFS or by combining it with other models, like capitation. In this context, our findings suggest that the regulator could make use of the described substitution pattern. Instead of purely setting prices based on the cost of providing a service, the regulator could increase fees for high-value services and thus incentivize physicians to provide more of them. Equivalently, fees for services that are of low medical value could be reduced (see Chandra et al., 2011; Chernew et al., 2007; Schwartz et al., 2014). Implementing such policies is in line with the recent wave of educational campaigns that propose a reduction of clinically unnecessary medical tests, treatments, and procedures.[51] Thus, implementing a reformed system that exploits the substitution responses described in our analysis is a promising avenue to reduce inefficiencies in healthcare provision. Because we find only a negligible effect of the substitution channel on total health care costs, our results suggest that such a reform may be cost-neutral in the short run.

3.7 Conclusion

This paper addresses physician responses to an exogenous and large-scale revision of medical service reimbursement. As shown in previous work, there are two potential types of responses: shifts in aggregate healthcare

[51] `www.choosingwisely.org`. Similar campaigns have been launched in Canada, the Netherlands, England, Japan, Australia, New Zealand, Germany, Italy, Switzerland, Wales and Denmark.

supply and substitutions between services. The present analysis is able to isolate the two channels and provides evidence on the relevance of both. More precisely, our results contribute in four ways. First, they suggest that physicians increase their supply when their aggregate revenue drops, thereby confirming existing findings. Further, physicians substitute from relatively less to more attractive services. Third, substituting between services is only relevant for physicians whose revenue was not reduced substantially. In cases with a marked revenue loss, physicians increase aggregate supply and thus are likely to seek an income target. This behavior is consistent with physician-induced demand. Finally, in terms of overall costs, the adjustment in aggregate supply is the main driver of changes in revenue.

Concerning the policy itself, our calculations suggest that the saving target of 0.47 billion was achieved in the first post-reform year. Our results on the substitution behavior suggest that adapting the fee structure may have (unintended) consequences for the service mix provided by physicians. Potentially, such effects may be adverse, for instance, when patients are undertreated or receive adverse overtreatment. Understanding the role of limitations posed on specific services constitutes a promising avenue for further research.

3.A Calculation of Relative Value Units

This section describes the calculation of the *TARMED*'s relative value units in more detail. The relative value units of a service are the sum of two components that account for the medical and the technical component of providing a service. Equation (3.A.1) illustrates this. The two components are described in the following.

$$\text{relative value units}_{s,i} = \text{MS}_{s,i} + \text{TS}_{s,i} \tag{3.A.1}$$

Medical component of relative value unit. The medical component MS_s reimburses medical work with the patient. The relative value units for the medical component is determined by six parameters.

$$MS_s = \frac{reference\ income \times dignity\ factor_i}{annual\ work\ load \times productivity_s} \times Scaling_i \times Minutage_s^M + (assistance_s) \tag{3.A.2}$$

- reference income = CHF 207,000
- dignity factor 0.905-2.2625 (pre-reform), 0.985 (post-reform), reimburses education and qualification, depends on specialty
- annual workload = 115,200 minutes
- productivity 45%-100%, depends on service (more specifically on the functional unit, *Sparte*, e.g. doctors office, types of OP).[52] The parameter measures the share of billable time. Basically: the time spent with patients as a percentage of total annual work time. Example: 0.85 for consultations.
- Scaling Factor. Adjusts for the different training durations. Scaling factor of 1 for all specialties except 0.93 for medical practitioner (*praktischer Arzt*). (Before 2018: 1 for all specialties) (shorter).
- medical assistance: only where applicable, depends on service
- Minutage is the (normative) time necessary to provide the service. Calculations based on cost models (GRAT, KOREG, ROKO for physician practices, INFRA for hospitals)

[52] A functional unit is an area of a practice where a certain bundle of services is provided. Each unit is characterized by a specific infrastructure (space, fixed and mobile equipment) and by a certain number of non-medical staff.

Technical component of relative value unit. The technical component TS_s reimburses the infrastructure and practice expenses that are necessary for providing a certain service. Relative value units are based on two parameters.

$$TS_s = cost\ per\ minute_s \times Minutage_s^T \tag{3.A.3}$$

- Costs per minute account for direct and indirect costs for infrastructure, material, wages (non-medical staff) etc. per minute. The numbers are based on cost models (GRAT, KOREG, ROKO for physician practices, INFRA for hospitals). The costs per minute vary by functional unit (*Sparte*), e.g. operation, consultation in doctor's office etc.

- Minutage measures the necessary time to provide a service. Numbers are determined in the same cost models.

3.B Compositional Changes

The *TARMED* reform changed the composition of a number of services. We summarize these *compositional changes* by the five categories described below. Each type is illustrated by a specific example and by the respective revenue share.

(1) Splitting services into finer, comparable sets. This intervention splits a given service into a set of new services, targeted at a specific patient population (e.g. children, elderly). The most important service in this category is *"Konsultation, jede weiteren 5 Min."* (service ID 00.0020). Starting from 2018Q1, physicians have to indicate the age group of their patient. This intervention forces physicians to report in more detail. As this is rather affecting the administrative burden and not the nature of the service it seems reasonable that these services are comparable over time. Thus, services falling into this category are included in our estimations. These services account for 14.1% of aggregate revenue between 2015Q4 and 2018Q4.

(2) Splitting services into finer, non-comparable sets. This category contains only one service, *"Glaskörperbiopsie für zytologische Diagnostik u/o intravitreale Injektion"* (service ID 08.3350). This service was divided into two new services both taking care of one part of the original biopsy. Thus, the sum of the new services cannot be compared to the old service. This category accounts for 0.5% of aggregate revenue between 2015Q4 and 2018Q4.

(3) Changing the Minutage. For a number of services, the *actual* time necessary to provide the service can be charged in fixed time intervals. One intervention changed the length of these intervals. Most prominently, this intervention affected the service *"Ärztliche Leistung in Abwesenheit des Patienten (inkl. Aktenstudium), pro 5 Min."* (service ID 00.0140). This used to be a rather vague service compensating the time a physician worked in the absence of the patient. The federal government decided to split the service into 24 new, and finer, services, as well as to change the *Minutage* interval from five to one minute. As a result, physicians need to report more precisely. It is reasonable to assume that this change aims at holding the physicians (more) accountable for what they charge when patients are not present. The 24 new categories are hardly comparable to the old service. Thus, these types of services are excluded from the regressions. This category accounts for 6.4% of aggregate revenue between 2015Q4 and 2018Q4.

(4) Introduction of Minutage. This measure aimed at how services are charged. Starting from 2018Q1 physicians have to report how long it took them to provide the services affected by this change. Most prominently, this concerns "*Kleine Untersuchung durch den Facharzt für Grundversorgung*" (service ID 00.0410). (Additionally, this service was also split into a finer set. Therefore, it is also affected by the first type of changes, discussed above.) The introduction of *Minutage* changes the nature of the services and thus makes the comparability over time infeasible. These services account for 5.6% of aggregate revenue between 2015Q4 and 2018Q4.

(5) Changing restrictions. This category contains only one service, "*Instruktion von Selbstmessungen, Selbstbehandlungen durch den Facharzt, pro 5 Min.*" (service ID 00.0610). For this service, an additional limitation was introduced: The service can be charged at most three times per visit. Including this service into our estimation would result in biased estimates as physicians are potentially restricted regarding their substitution behavior, post-reform. (Additionally, this service was also split into a finer set. Therefore, it is also affected by the first type of changes, discussed above.) This category accounts for 0.6% of aggregate revenue between 2015Q4 and 2018Q4.

3.C Assessing the Importance of the Substitution and the Volume Expansion Channel

The two models outlined in Section 3.5 identify the two types of potential responses to changes in prices: the substitution channel and the volume expansion channel. Because the models differ with respect to their observation unit and their outcome variable, we cannot directly compare them to understand which channel is more important in terms of overall costs. To serve this purpose, this section introduces a revenue decomposition similar to Equation (3.5.4). In essence, we decompose the change in log revenue for physician i from year $y-1$ to year y into three components: the change in revenue due to (i) a modified service mix (substitution), (ii) a new level of labor supply (volume expansion), and (iii) the change in fees (purely mechanical). Equation (3.C.1) formalizes this decomposition.

$$\begin{aligned}\ln R_{i,y} - \ln R_{i,y-1} = &\underbrace{\ln\left(\sum_s^S P_{s,y} \cdot Q_{i,s,y}\right) - \ln\left(\sum_s^S P_{s,y} \cdot Q_{i,s,y-1}\right)}_{\text{behavioral change}} \\ &+ \underbrace{\ln\left(\sum_s^S P_{s,y} \cdot Q_{i,s,y-1}\right) - \ln\left(\sum_s^S P_{s,y-1} \cdot Q_{i,s,y-1}\right)}_{\text{mechanical change}}\end{aligned} \tag{3.C.1}$$

$$\begin{aligned}\Leftrightarrow \ln R_{i,y} - \ln R_{i,y-1} = &\underbrace{\ln\left(\sum_s^S P_{s,y} \cdot Q_{i,s,y}\right) - \ln\left(\sum_s^S P_{s,y} \cdot \tilde{Q}_{i,s,y}\right)}_{\text{substitution channel}} \\ &+ \underbrace{\ln\left(\sum_s^S P_{s,y} \tilde{Q}_{i,s,y}\right) - \ln\left(\sum_s^S P_{s,y} \cdot Q_{i,s,y-1}\right)}_{\text{volume expansion channel}} \\ &+ \underbrace{\ln\left(\sum_s^S P_{s,y} \cdot Q_{i,s,y-1}\right) - \ln\left(\sum_s^S P_{s,y-1} \cdot Q_{i,s,y-1}\right)}_{\text{mechanical change}}\end{aligned} \tag{3.C.2}$$

$R_{i,y}$ measures physician i's revenue in year y, due to charging volume $Q_{i,s,y}$ of service s at fee $P_{s,y}$. $\tilde{Q}_{i,s,y}$ is a counterfactual volume, adjusted for the growth in work hours (see description below). This model is more general

than Equation (3.5.4) because it is defined for all years. In contrast, equation (3.5.4) is only defined for the revenue change from 2017 to 2018.
Note that we have introduced two counterfactual revenues. First, $\sum_s^S P_{i,s,y} \cdot Q_{i,s,y-1}$ measures the (artificial) revenue in year y, holding all service volumes fixed at prior-year levels. Second, $\sum_s^S P_{i,s,y}\tilde{Q}_{i,s,y}$ constitutes the revenue in year y holding physician's service mix[53] fixed at the prior-year level (adjusted for the average growth of her labor supply). In an intermediate step, $\tilde{Q}_{i,s,y}$ multiplies the previous year's volume by the growth rate of a physician's annual working hours WH_i, i.e.:[54]

$$\tilde{Q}_{i,s,y} = Q_{i,s,y-1} \cdot \frac{WH_{i,y}}{WH_{i,y-1}}.$$

The fee schedule provides a time estimate for 90% of all services, allowing us to reconstruct the annual working hours. The measure is highly correlated with revenue (correlation of 0.98), arguably making working hours a good measure of overall supply.

In terms of interpretation, the *mechanical change* inherits the intuition of the Revenue Decomposition Model in Equation (3.5.4): It measures by how much revenue changes due to the fee adjustments, holding the supply level fixed.[55] Intuitively, the *volume expansion channel* reflects how a change in total working hours (as a proxy for total supply) translates to revenue (holding the service mix fixed). Finally, the *substitution channel* measures the difference in revenue due to changes in the physician's service mix. If a physician adapts the volume of all services by exactly the same factor, the *substitution channel* will always be zero.[56] In other words, the *substitution channel* captures all volume changes not stemming from an overall shift in labor supply.

[53]Let $share_s = \frac{Q_s}{\sum_s Q_s}$ measure the share of service s in physicians i's practice in year y. We define the service mix as the set of $share_s \forall s$. It can be interpreted as the share of labor physician i allocates to service s. Note that the volume shares of all services add to 1, $\sum_s \frac{Q_s}{\sum_s Q_s}$ and give us a notion on the composition of a physician's supplied services.

[54]Note that we can not use a revenue-based growth measure as this would be affected by the new fees.

[55]Note that the definition differs somewhat from the definition used earlier in the paper (see the Revenue Decomposition Model in Subsection 3.5.2. Specifically, Equation (3.C.1) does not adjust for different trends in volumes across physicians and services. The reason is

[56]Note that this is a rather rough way to conceptualize an aggregate measure of how strongly physicians substitute. Nevertheless, a quantifiable measure is sufficient to compare substitution to aggregate supply changes.

Figure 3.7 in Appendix 3.F shows the overall change in revenue from 2017 to 2018, the *volume expansion channel* and the *substitution channel* by the deciles of the *mechanical change*. First, Subfigure (a) shows how physician revenue changed from 2017 to 2018. This corresponds to the left hand side of Equation (3.C.1). The revenue change is shown as a function of the mechanical change (also according to Equation (3.C.1)).[57] The subfigure again illustrates that physicians did in fact respond to the fee reform. Absent any behavior changes, the solid line would coincide with the 45-degree line, because the mechanical change would translate one-by-one to (realized) revenue. Except for the mechanical changes above 2%, revenue change lies above the 45-degree line, thus, physicians in these groups successfully mitigate the income losses due to the reform.
Subfigure (b) shows the three parts that compose total revenue change in Equation (3.C.1). The results suggest that mechanical revenue losses are mitigated through the volume expansion channel, especially at the lower end of the mechanical change distribution. A 5% (mechanical) revenue loss increases supply by roughly 4%. In contrast, the substitution channel's aggregate effect is close to zero and invariant to the *mechanical change*. We conclude that only the aggregate supply channel is relevant in terms of total healthcare costs.

Does this result imply that the substitution channel is negligible? Not necessarily. The substitution patterns described in Subsection 3.6.1 do not manifest in increased *costs* in the short term. However, it may be that the *quality* of healthcare provision is affected. It is a priori unclear in which direction. The substitution channel may provide a possibly cost-neutral opportunity to improve healthcare quality by increasing the relative financial attractiveness of high-quality procedures (see Chandra et al., 2011; Chernew et al., 2007; Schwartz et al., 2014). Analyzing this avenue is left to future research.

[57]Note that while this is an exact decomposition, we preferred to estimate a fractional polynomial regression with degree 2 to approximate the relation.

3.D Prediction of Physican-Specific Service Volumes

We predict physician-specific service volume based on pre-reform observations. In a first step, the aggregate average change in quantity was computed for service s and physician speciality g. In a second step, a percentage prediction for every service and physician speciality pair was computed. Finally, it is assumed that physicians within the same speciality share the same percentage changes of their quantities.

$$\text{Step 1:} \quad \overline{\Delta Q}_{g,s,j} = \frac{1}{T/4} \sum_{j}^{T/4} \Delta Q_{g,s,j+4k} \qquad j \in \{1,2,3,4\}$$

$$\text{Step 2:} \quad \%\hat{Q}_{g,s,j_{2018}} = \left(Q_{g,s,j_{2018}-4} + \overline{\Delta Q}_{g,s,j}\right) / Q_{g,s,j_{2018}-4}$$

$$\text{Step 3:} \quad \hat{Q}_{i,s,j_{2018}} = \left(1 + Q_{i,s,j_{2018}-4}\right) \%\hat{Q}_{g,s,j_{2018}}$$

3.E Additional Tables

Table 3.2: Descriptive Statistics: Quarterly BITE by Medical Specialty (Top 12)

	Mean	SD	Median	IQR	Min	Max	N
Losers							
Radiologists	−0.111	0.056	−0.129	0.057	−0.210	0.019	430
Gastroenterologists	−0.100	0.034	−0.099	0.036	−0.222	0.030	668
Cardiologists	−0.090	0.035	−0.094	0.047	−0.216	0.038	1,270
Psychiatrists	−0.030	0.010	−0.036	0.014	−0.093	0.048	7,275
Total	−0.046	0.036	−0.037	0.010	−0.222	0.048	9,643
Midrange							
OB-GYNs	0.019	0.012	0.020	0.011	−0.094	0.046	3,718
Dermatologists	0.017	0.022	0.023	0.028	−0.101	0.045	1,317
Group practices	0.009	0.059	0.029	0.032	−0.472	0.051	1,112
Oto-Rhino Laryngologists	0.024	0.018	0.030	0.016	−0.096	0.043	1,114
Total	0.018	0.027	0.022	0.017	−0.472	0.051	7,261
Winners							
Medical Practitioners	0.031	0.011	0.032	0.011	−0.094	0.052	2,698
Ophthalmologists	0.006	0.075	0.034	0.007	−0.473	0.042	2,319
Pediatricians	0.040	0.013	0.043	0.006	−0.095	0.051	3,059
GPs	0.043	0.011	0.046	0.005	−0.130	0.095	15,609
Total	0.038	0.028	0.044	0.010	−0.473	0.095	23,685

Notes: The table shows descriptive statistics for the quarterly BITE variable by medical specialty for the estimation sample. N is physicians $\times$ quarter.

Table 3.3: Descriptive Statistics: BITE by Medical Specialty (Non-Top 12)

	Mean	SD	Median	IQR	Min	Max	N
Plastic/Reconstructive Surgeons	−0.045	0.050	−0.053	0.083	−0.132	0.044	228
Neurologists	−0.043	0.033	−0.045	0.050	−0.148	0.048	755
Angiologists	−0.035	0.012	−0.034	0.014	−0.070	−0.000	287
Child psychiatrists	−0.026	0.012	−0.027	0.019	−0.063	0.047	986
Pneumologists	−0.025	0.031	−0.026	0.038	−0.102	0.049	516
Endocrinologists	−0.021	0.024	−0.020	0.031	−0.118	0.047	355
Rheumatologists	−0.014	0.026	−0.017	0.034	−0.085	0.048	926
Other	−0.015	0.052	−0.010	0.071	−0.183	0.051	344
Allergologists	−0.009	0.025	−0.010	0.031	−0.074	0.042	240
Medical Oncologists	−0.000	0.018	−0.003	0.022	−0.065	0.049	349
Anaesthesiologists	−0.000	0.043	−0.000	0.070	−0.139	0.053	258
Surgeons	−0.006	0.042	−0.000	0.062	−0.157	0.247	677
Hematologists	−0.001	0.034	0.001	0.048	−0.093	0.048	109
Urologists	−0.001	0.020	0.003	0.021	−0.101	0.037	616
Neurosurgeons	−0.008	0.042	0.006	0.064	−0.149	0.049	169
Nephrologists	0.005	0.025	0.008	0.016	−0.091	0.042	101
Physical Medicine and Rehabilitation	0.006	0.026	0.008	0.032	−0.058	0.046	174
Orthopaedic Surgeons	0.013	0.034	0.023	0.032	−0.134	0.370	1,396
Total	−0.012	0.036	−0.013	0.049	−0.183	0.370	8,486

Notes: The table shows descriptive statistics for the quarterly BITE variable by medical specialty for the estimation sample. "Other" contains specialties with 15 or fewer observations per quarter. N is physicians × quarter.

Table 3.4: Substitution Model: Regression Estimates

	Δ lnQ		
	(1) Linear	(2) Kink	(3) Quadratic
$\Delta \ln P$	0.253***	0.979***	0.368***
	(0.013)	(0.031)	(0.020)
$1(\Delta lnP < 0)\Delta \ln P$		−0.894***	
		(0.035)	
$(\Delta lnP)^2$			0.174***
			(0.021)
R^2	0.095	0.090	0.095
adj. R^2	0.090	0.085	0.090
N Phys × Quarter	155,481	155,481	155,481
N	28,687,618	28,687,618	28,687,618

Notes: The table shows estimates for the substitution model. The outcome variable is $\Delta \ln Q$ in all columns. All models additionally include a constant, service fixed effects, and physician-by-quarter fixed effects. Observations are weighted with the prior-year quarter log revenue. Standard errors are clustered at the physician-and-quarter level in parentheses, *** $p < 0.001$ ** $p < 0.01$, * $p < 0.05$.

Table 3.5: Substitution Model: Regression Estimates II

	(1) $\Delta \ln Q$	
$\Delta \ln P$	0.452***	(0.048)
× Dermatologists	0.413***	(0.124)
× OB-GYN	−0.122	(0.072)
× Psychiatrists	−1.660***	(0.181)
× Ophthalmologists	0.133	(0.104)
× Oto-Rhino Laryngologists	−0.072	(0.154)
× Pediatricians	0.233*	(0.110)
× Radiologists	−0.498***	(0.115)
× GPs	0.149**	(0.058)
× Cardiologists	0.593***	(0.133)
× Gastroenterologists	−0.478***	(0.101)
× Medical Practitioners	−0.215	(0.126)
× Group practices	−0.130	(0.086)
$(\Delta \ln P)^2$	0.503***	(0.149)
× Dermatologists	2.305**	(0.847)
× OB-GYN	0.806	(0.469)
× Psychiatrists	−5.404**	(1.960)
× Ophthalmologists	−0.239	(0.162)
× Oto-Rhino Laryngologists	1.631	(1.055)
× Pediatricians	−1.047	(1.231)
× Radiologists	−0.401	(0.398)
× GPs	−0.085	(0.167)
× Cardiologists	1.354***	(0.405)
× Gastroenterologists	−0.617***	(0.170)
× Medical Practitioners	−0.487**	(0.177)
× Group practices	−0.440**	(0.169)
R^2	0.096	
adj. R^2	0.091	
N Phys × Quarter	155,481	
N	28,687,618	

Notes: The table shows estimates for the substitution model (quadratic fit) where the fee parameters vary for the twelve largest medical specialties. The outcome variable is $\Delta \ln Q$ in all columns. All models additionally include a constant, service fixed effects, and physician-by-quarter fixed effects. Observations are weighted with the prior-year quarter log revenue. Standard errors clustered at the physician-and-quarter level in parentheses,
*** $p < 0.001$ ** $p < 0.01$, * $p < 0.05$.

Table 3.6: Total Healthcare Supply: Regression Estimates (Consultations)

	Δ ln Consultation	
	(1) Baseline	(2) Covariates
BITE (quarterly)	1.090***	0.935***
	(0.009)	(0.009)
BITE × Radiologists	−1.345***	−1.205***
	(0.028)	(0.027)
BITE × Cardiologists	−2.171***	−2.006***
	(0.022)	(0.022)
BITE × Gastroenterologists	−2.601***	−2.445***
	(0.027)	(0.026)
Covariates (Prior Year Quarter):		
Limited Services(Revenue Share)		−0.425***
		(0.013)
# Distinct Services		−0.005***
		(0.000)
Female (Revenue Share)		0.203***
		(0.013)
Young (Revenue Share)		−0.194***
		(0.007)
Morbid (Revenue Share)		−0.056***
		(0.010)
Low Deductible (Revenue Share)		−0.030***
		(0.005)
Physician Density		0.365***
		(0.007)
Constant	−0.005***	−0.421***
	(0.000)	(0.023)
R^2	0.254	0.267
adj. R^2	0.246	0.259
N Physician	9,973	9,973
N Physician × Quarter	870,354	870,354

Notes: The table shows estimates for the consultation model. Observations are weighted with the prior-year quarter log revenue. All models include physician fixed-effects. The covariates in column (2) measure prior-year-quarter levels of the number of distinct services, the cantonal physician density, and the revenue share of services that face some limitations. Further, the models account for differences in the patient groups by including the revenue shares of the following groups: women, young (below age 36), morbid (hospital stay in the previous year), low deductible (CHF 300). Standard errors clustered at the physician-and-quarter level in parentheses, *** $p < 0.001$ ** $p < 0.01$, * $p < 0.05$.

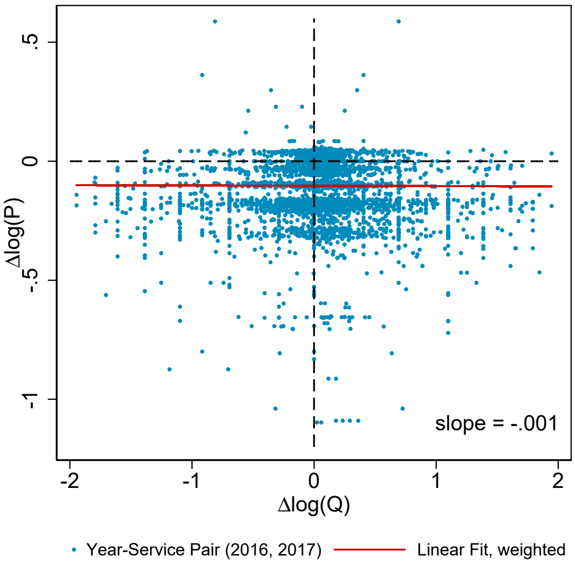
Table 3.7: Total Health Care Supply: Regression Estimates (Revenue Decomposition)

	Behavioral Change	
	(1) Baseline	(2) Covariates
Mechanical Change	−0.008	−0.168***
	(0.031)	(0.039)
Mechanical Change × Psychiatrists	−1.080***	−1.284***
	(0.282)	(0.295)
Mechanical Change × Radiologists	0.051	0.118
	(0.181)	(0.191)
Mechanical Change × Cardiologists	−0.511*	−0.087
	(0.222)	(0.229)
Mechanical Change × Gastroenterologists	−0.831**	−0.508
	(0.255)	(0.263)
Limited Services (Revenue Share)		−0.076***
		(0.013)
# Distinct Services		−0.000
		(0.000)
Female (Revenue Share)		0.013
		(0.012)
Young (Revenue Share)		0.066***
		(0.008)
Morbid (Revenue Share)		0.133**
		(0.043)
Low Deductible (Revenue Share)		0.015
		(0.014)
Physician Density		0.009***
		(0.001)
Psychiatrists	−0.034***	−0.052***
	(0.009)	(0.010)
Radiologists	0.013	0.030
	(0.024)	(0.025)
Cardiologists	−0.021	−0.009
	(0.024)	(0.025)
Gastroenterologists	−0.054	−0.047
	(0.030)	(0.031)
Constant	−0.036***	−0.034
	(0.001)	(0.018)
N	13,881	13,771

Notes: The table shows median regression estimates for the revenue decomposition model. Observations are weighted with the prior-year log revenue. The covariates in column (2) measure prior-year levels of the number of distinct services, the cantonal physician density, and the revenue share of services that face some limitations. Further, column (2) accounts for differences in the patient groups by including the revenue shares of the following groups: women, young (below age 36), morbid (hospital stay in the previous year), low deductible (CHF 300). Standard errors in parentheses, *** $p < 0.001$ ** $p < 0.01$, * $p < 0.05$.

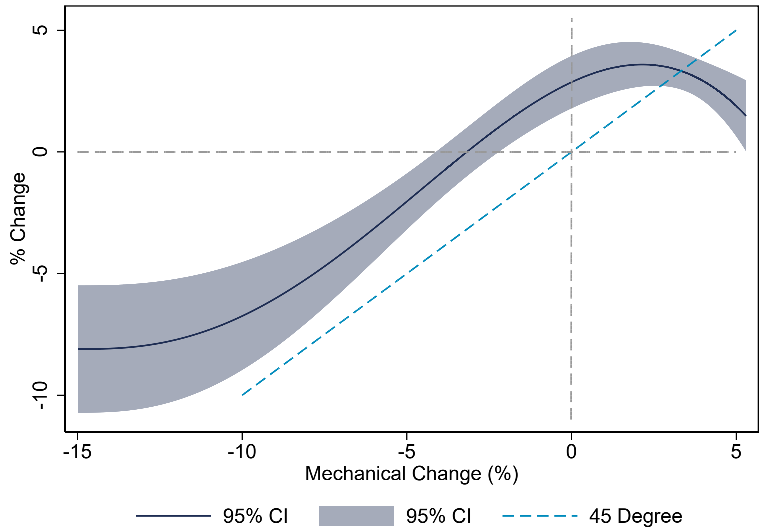
3.F Additional Figures

Figure 3.6: Pre-Treatment Changes in Service Volumes

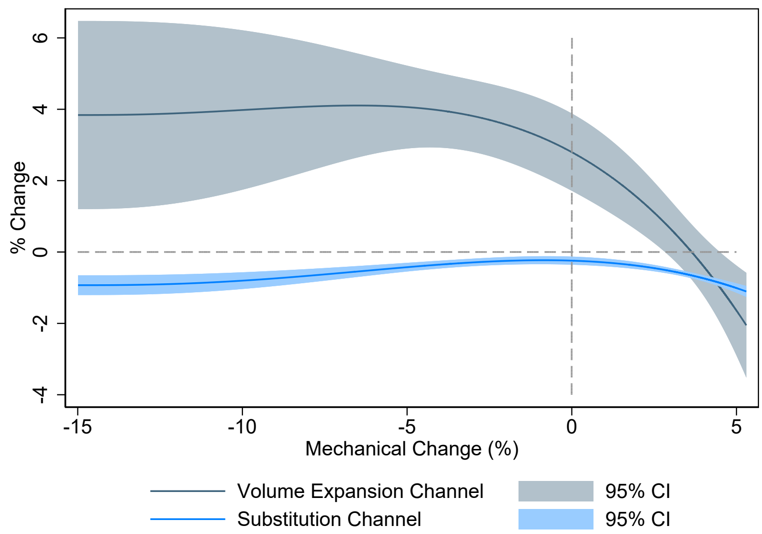


Note: The figure shows pre-reform changes in service volumes and relative fee changes for each service. The linear fit uses prior-year revenue weights.

Figure 3.7: Decomposition of Changes in Revenue, 2017 to 2018



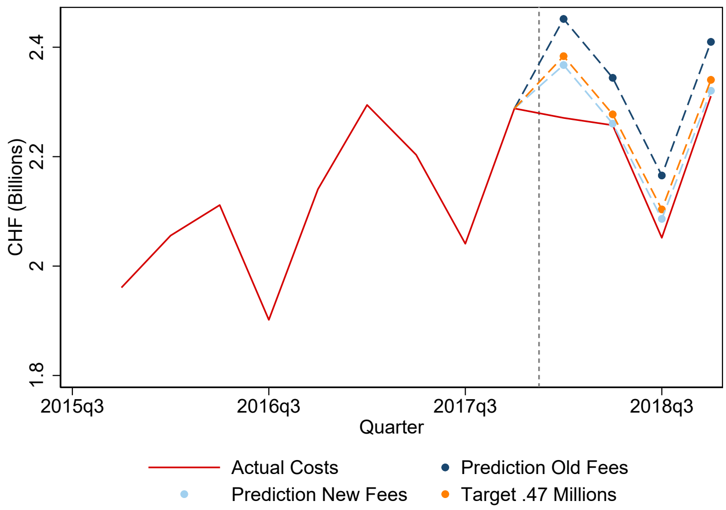
(a) Total Revenue



(b) Revenue Decomposition

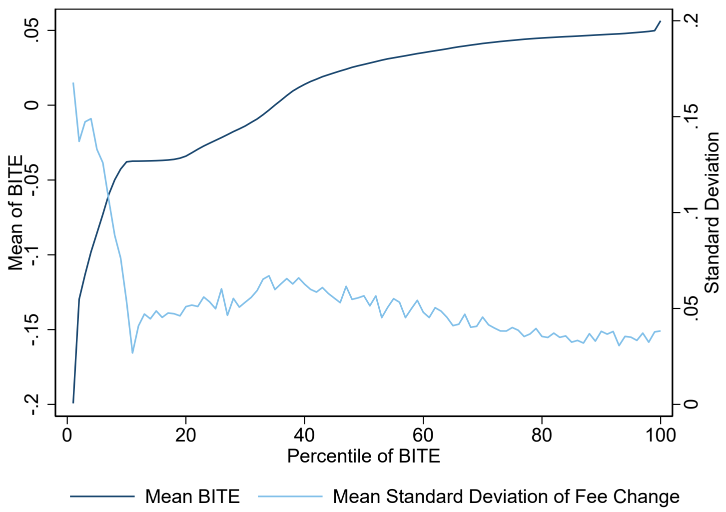
Note: The figure illustrates how physician revenue changed from 2017 to 2018 and relates it to the mechanical revenue change due to the reform. The solid lines in both subfigures show (counterfactual) revenue changes by values of mechanical change estimated by a fractional polynomial regression of degree 2. Subfigure (a) shows the mean revenue change from 2017 to 2018, corresponding to the left hand side of Equation (3.C.1), by decile of mechanical change. Subfigure (b) illustrates the decomposed revenue changes according to Equation (3.C.1).

Figure 3.8: Total and Predicted Costs, Estimation Sample (Quarterly)



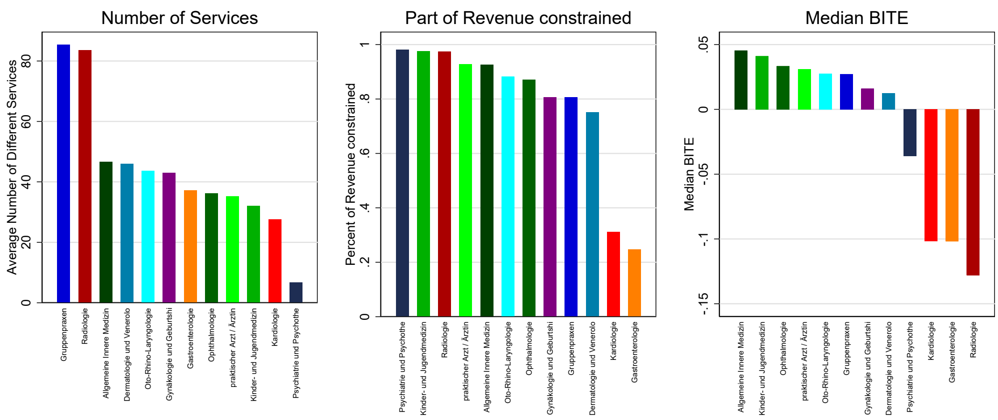
Note: The figure shows aggregate quarterly revenue for the outpatient sector covered by MHI. The dashed vertical lines mark the *TARMED* reforms. Data is based on the estimation sample.

Figure 3.9: Distribution and Standard Deviation of BITE



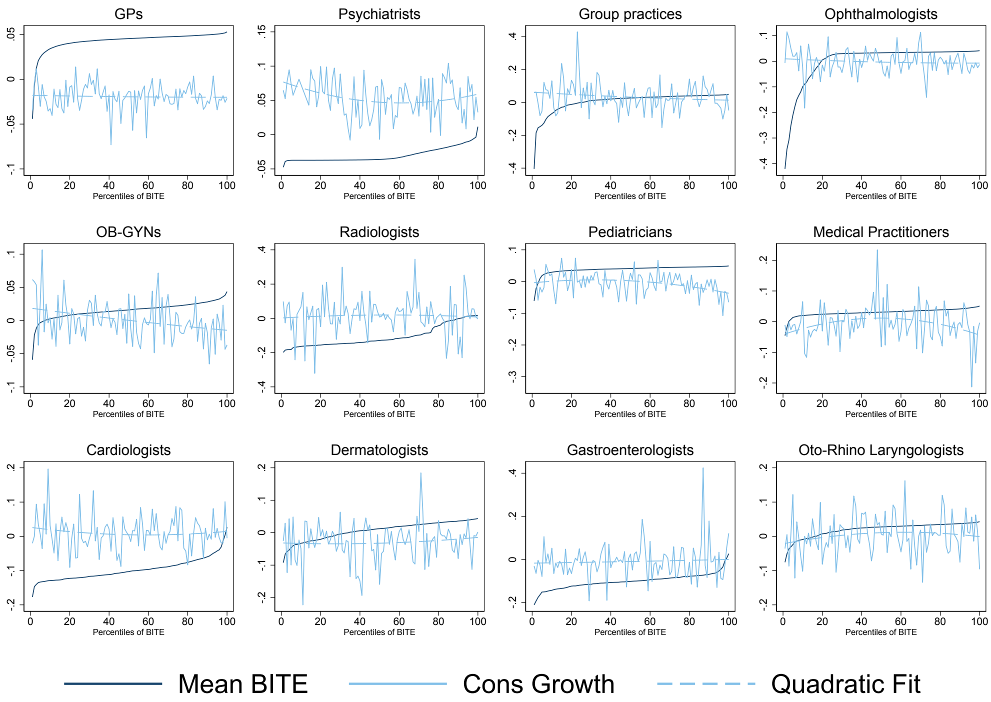
Note: The figure shows the average *BITE* by the percentiles of *BITE* (dark blue line). The light blue line is the standard deviation of *BITE* within each percentile.

Figure 3.10: Statistics for the Top 12 specialties



Note: This figure is based on the estimation sample. The color scheme follows Figure 3.4

Figure 3.11: Pre-reform Consultation Growth by BITE



Note: The figure shows average BITE, consultation growth and the quadratic fit by percentiles of BITE for the 12 largest specialties for the pre-reform period.

Bibliography

Aalen, Odd, "Statistical inference for a family of counting process." PhD dissertation, Ph. D. thesis, University of California, Berkeley 1975.

— , "A model for nonparametric regression analysis of counting processes," in "Mathematical statistics and probability theory," Springer, 1980, pp. 1–25.

— **, Ornulf Borgan, and Hakon Gjessing**, *Survival and event history analysis: a process point of view*, Springer Science & Business Media, 2008.

Alessi, Lucia and Mark Kerssenfischer, "The response of asset prices to monetary policy shocks: Stronger than thought," *Journal of Applied Econometrics*, apr 2019, *34* (5), 661–672.

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*, 2011, *35* (1), 130–141.

Anderson, Brooke G. and Michelle L. Bell, "Weather-related mortality: how heat, cold, and heat waves affect mortality in the United States," *Epidemiology (Cambridge, Mass.)*, 2009, *20* (2), 205.

Arrow, Kenneth J., "Uncertainty and the Welfare Economics of Medical Care," *The American Economic Review*, 1963, *53* (5), 941–973.

Ashraf, Nava, "Spousal Control and Intra-Household Decision Making: An Experimental Study in the Philippines," *American Economic Review*, 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*, 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*, 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*, 2017, p. j14.

Bennett, Steve, "Log-logistic regression models for survival data," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 1983, *32* (2), 165–171.

Beran, Rudolf, "Nonparametric regression with randomly censored survival data," 1981.

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*, 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*, 2016, *106* (2), 387–435.

__ **,** __ **, and** __ , "Children, Time Allocation, and Consumption Insurance," *Journal of Political Economy*, 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?," Technical Report, National Bureau of Economic Research 2019.

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

Brekke, Kurt R., Tor Helge Holmås, Karin Monstad, and Odd Rune Straume, "Do treatment decisions depend on physicians' financial incentives?," *Journal of Public Economics*, 2017, *155*, 74–92.

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

Buckley, Jonathan and Ian James, "Linear regression with censored data," *Biometrika*, 1979, *66* (3), 429–436.

Burkhard, Daniel, Christian P. R. Schmid, and Kaspar Wüthrich, "Financial incentives and physician prescription behavior: Evidence from dispensing regulations," *Health Economics*, 2019, *28* (9), 1114–1129.

Caballero, Ricardo J and Alp Simsek, *Asset prices and aggregate demand in a" Covid-19" shock: A model of endogenous risk intolerance and LSAPS*, National Bureau of Economic Research, 2020.

Canadian Press, "Trudeau and Scheer blame one another for blocking emergency disability benefit," *The Canadian Press*, 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*, 2017, *107* (12), 3917–3946.

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

Chandra, Amitabh, David Cutler, and Zirui Song, "Chapter Six - Who Ordered That? The Economics of Treatment Choices in Medical Care," in Mark V. Pauly, Thomas G. Mcguire, and Pedro P. Barros, eds., *Handbook of Health Economics*, Vol. 2 of *Handbook of Health Economics*, Elsevier, 2011, pp. 397 – 432.

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

Chernew, Michael E., Allison B. Rosen, and A. Mark Fendrick, "Value-Based Insurance Design: By abandoning the archaic principle that all services must cost the same for all patients, we can move to a high-value health system.," *Health Affairs*, 2007, *26* (Suppl2), 195–203.

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.

— **,** — **, 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.

— **,** — **, Blaise Melly, and Kaspar Wüthrich**, "Generic inference on quantile and quantile effect functions for discrete outcomes," *Journal of the American Statistical Association*, 2019.

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

Chetty, Raj, "A general formula for the optimal level of social insurance," *Journal of Public Economics*, 2006, *90* (10-11), 1879–1901.

— , "Moral hazard versus liquidity and optimal unemployment insurance," *Journal of political Economy*, 2008, *116* (2), 173–234.

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

Clemens, Jeffrey and Joshua D Gottlieb, "Do physicians' financial incentives affect medical treatment and patient health?," *American Economic Review*, 2014, *104* (4), 1320–49.

— **and Joshua D. Gottlieb**, "In the shadow of a giant: Medicare's influence on private physician payments," *Journal of Political Economy*, 2017, *125* (1), 1–39.

Codespote, Suzanne M., William J. London, and John D. Shatto, "Physician Volume & Intensity Response, Memorandum to Richard S. Foster, Office of the Actuary," *Health Care Financing Administration August*, 1998, *13*.

Congressional Budget Office, "Factors Underlying the Growth in Medicare's Spending for Physicians' Services," Background Paper 2597, The Congress of the United States 2007.

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

Cromwell, Jerry and Janet B Mitchell, "Physician-induced demand for surgery," *Journal of health economics*, 1986, *5* (4), 293–313.

Currie, Janet, Wanchuan Lin, and Wei Zhang, "Patient knowledge and antibiotic abuse: Evidence from an audit study in China," *Journal of Health Economics*, 2011, *30* (5), 933 – 949.

Dabrowska, Dorota M, "Non-parametric regression with censored survival time data," *Scandinavian Journal of Statistics*, 1987, pp. 181–197.

_ , "Uniform consistency of the kernel conditional Kaplan-Meier estimate," *The Annals of Statistics*, 1989, pp. 1157–1167.

Dahl, Gordon B., Andreas Ravndal Kostøl, and Magne Mogstad, "Family Welfare Cultures," *The Quarterly Journal of Economics*, 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*, 2016, *22* (1), 302–324.

Delgado, Miguel A, Andrés García-Suaza, and Pedro HC Sant'Anna, "Distribution Regression in Duration Analysis: an Application to Unemployment Spells," *arXiv preprint arXiv:1904.06185*, 2019.

der Vaart, Aad W. 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*, 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.

Doksum, Kjell A and Miriam Gasko, "On a correspondence between models in binary regression analysis and in survival analysis," *International Statistical Review/Revue Internationale de Statistique*, 1990, pp. 243–252.

Domenighetti, Gianfranco, Antoine Casabianca, Felix Gutzwiller, and Sebastiano Martinoli, "Revisiting the most informed consumer of surgical services: the physician-patient," *International journal of technology assessment in health care*, 1993, *9* (4), 505–513.

Dowell, Scott F. and Mei Shang Ho, "Seasonality of infectious diseases and severe acute respiratory syndrome - what we don't know can hurt us," *The Lancet Infectious Diseases*, 2004, *4* (11), 704 – 708.

Dranove, David and Paul Wehner, "Physician-induced demand for childbirths," *Journal of health economics*, 1994, *13* (1), 61–73.

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

Dunn, Abe and Adam Hale Shapiro, "Physician competition and the provision of care: evidence from heart attacks," *American Journal of Health Economics*, 2018, *4* (2), 226–261.

Efron, Bradley, "The two sample problem with censored data," in "Proceedings of the fifth Berkeley symposium on mathematical statistics and probability," Vol. 4 1967, pp. 831–853.

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

Enthoven, A., "Consumer-Choice Health Plan," *New England Journal of Medicine*, 1978, *298*, 650–658, 709–720.

Evans, Robert G, "Supplier-induced demand: some empirical evidence and implications," in "The economics of health and medical care," Springer, 1974, pp. 162–173.

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

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

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

Fleming, Thomas R and David P Harrington, *Counting processes and survival analysis*, Vol. 169, John Wiley & Sons, 2011.

FOPH, *Inhalt der Änderung und Kommentar Verordnung über die Festlegung und die Anpassung von Tarifstrukturen in der Krankenversicherung* Federal Office of Public Health 2017.

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*, 2014, *6* (2), 291–337.

Fuchs, Victor R., "The Supply of Surgeons and the Demand for Operations," *The Journal of Human Resources*, 1978, *13*, 35–56.

Fuwa, Makiko, "Macro-level Gender Inequality and the Division of Household Labor in 22 Countries," *American Sociological Review*, 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*, 2017, *9* (3), 229–261.

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

Gottschalk, Felix, Wanda Mimra, and Christian Waibel, "Health services as credence goods: A field experiment," *The Economic Journal*, 2020, *130* (629), 1346–1383.

Goussé, Marion, Nicolas Jacquemet, and Jean-Marc Robin, "Marriage, Labor Supply, and Home Production," *Econometrica*, 2017, *85* (6), 1873–1919.

Gruber, Jon, John Kim, and Dina Mayzlin, "Physician Fees and Procedure Intensity: The Case of Cesarean Delivery," *Journal of Health Economics*, 1999, *18* (4), 473–490.

Gruber, Jonathan and Maria Owings, "Physician Financial Incentives and Cesarean Section Delivery," *Rand Journal of Economics*, Spring 1996, *27* (1), 99–123.

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

Hao, Lingxin and Daniel Naiman, *Quantile Regression*, SAGE Publications, Inc., 2007.

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*, 2015, *31* (3), 603–610.

Hostettler, Stefanie and Esther Kraft, "FMH-Ärztestatistik 2017 - aktuelle Zahlen," *Schweizerische Ärztezeitung*, 2018, *99* (1314), 408–413.

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

Johnson, Erin M. and M. Marit Rehavi, "Physicians treating physicians: Information and incentives in childbirth," *American Economic Journal: Economic Policy*, 2016, *8* (1), 115–41.

Johnston, Andrew C and Alexandre Mas, "Potential unemployment insurance duration and labor supply: The individual and market-level response to a benefit cut," *Journal of Political Economy*, 2018, *126* (6), 2480–2522.

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

Kaplan, Edward L and Paul Meier, "Nonparametric estimation from incomplete observations," *Journal of the American statistical association*, 1958, *53* (282), 457–481.

Kaplan, Greg, Benjamin Moll, and Giovanni L Violante, "The great lockdown and the big stimulus: Tracing the pandemic possibility frontier for the US," Technical Report, National Bureau of Economic Research 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*, 2019, *11* (4), 181–209.

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

_ **and Olga Geling**, "Reappraising medfly longevity: a quantile regression survival analysis," *Journal of the American Statistical Association*, 2001, *96* (454), 458–468.

_ **and Yannis Bilias**, "Quantile regression for duration data: A reappraisal of the Pennsylvania reemployment bonus experiments," *Empirical Economics*, 2001, *26* (1), 199–220.

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

Krueger, Alan B and Andreas I Mueller, "A contribution to the empirics of reservation wages," *American Economic Journal: Economic Policy*, 2016, *8* (1), 142–79.

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*, 2011, *28* (5), 565–582.

Lalive, Rafael, "Unemployment benefits, unemployment duration, and post-unemployment jobs: A regression discontinuity approach," *American Economic Review*, 2007, *97* (2), 108–112.

_ **, Camille Landais, and Josef Zweimüller**, "Market externalities of large unemployment insurance extension programs," *The American Economic Review*, 2015, pp. 3564–3596.

Landais, Camille, "Assessing the welfare effects of unemployment benefits using the regression kink design," *American Economic Journal: Economic Policy*, 2015, *7* (4), 243–78.

Lee, Siha, "Spousal Labor Supply, Caregiving, and the Value of Disability Insurance," *Caregiving, and the Value of Disability Insurance (June 23, 2020)*, 2020.

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

London, Wayne P. and James A. Yorke, "Recurrent outbreaks of measles, chickenpox and mumps: I. Seasonal variation in contact rates," *American journal of epidemiology*, 1973, *98* (6), 453–468.

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

Marinescu, Ioana and Daphné Skandalis, "Unemployment insurance and job search behavior," *The Quarterly Journal of Economics*, 2021, *136* (2), 887–931.

McClellan, Mark, "Reforming payments to healthcare providers: The key to slowing healthcare cost growth while improving quality?," *Journal of Economic Perspectives*, 2011, *25* (2), 69–92.

McCullagh, Peter, "Regression models for ordinal data," *Journal of the Royal Statistical Society: Series B (Methodological)*, 1980, *42* (2), 109–127.

McGuire, Thomas G, "Physician agency," *Handbook of health economics*, 2000, *1*, 461–536.

_ **and Mark V Pauly**, "Physician response to fee changes with multiple payers," *Journal of health economics*, 1991, *10* (4), 385–410.

Melly, Blaise and Rafael Lalive, *Estimation, Inference, and Interpretation in the Regression Discontinuity Design*, Universität Bern, Faculty of Business, Economics and Social Sciences, 2020.

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 intrahousehold bargaining power," *Empirical Economics*, 2019.

Moineddin, Rahim, Jason X. Nie, Gabrielle Domb, Alan M. Leong, and Ross E. G. Upshur, "Seasonality of primary care utilization for respiratory diseases in Ontario: a time-series analysis," *BMC Health Services Research*, 2008, *8* (1), 160.

Mueller, Andreas I, Johannes Spinnewijn, and Giorgio Topa, "Job Seekers' Perceptions and Employment Prospects: Heterogeneity, Duration Dependence, and Bias," *American Economic Review*, 2021, *111* (1), 324–63.

OECD, "System of Health Accounts 1997-2017," Online 2019.

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

Peng, Limin, "Self-consistent estimation of censored quantile regression," *Journal of Multivariate Analysis*, 2012, *105* (1), 368–379.

_ **and Yijian Huang**, "Survival analysis with temporal covariate effects," *Biometrika*, 2007, *94* (3), 719–733.

_ **and** _ , "Survival analysis with quantile regression models," *Journal of the American Statistical Association*, 2008, *103* (482), 637–649.

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

Pietro, Carlo De, Paul Camenzind, Isabelle Sturny, Luca Crivelli, Suzanne Edwards-Garavoglia, Anne Spranger, Friedrich Wittenbecher, and Wilm Quentin, "Switzerland: health system review.," *Health Systems in Transition*, 2015, *17* (4), 1–288.

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.

Portnoy, Stephen, "Censored regression quantiles," *Journal of the American Statistical Association*, 2003, *98* (464), 1001–1012.

Powell, James L, "Censored regression quantiles," *Journal of econometrics*, 1986, *32* (1), 143–155.

R Core Team, *R: A Language and Environment for Statistical Computing* R Foundation for Statistical Computing 2020.

Reinhardt, Uwe E., "The Economist's Model of Physician Behavior," *JAMA*, 02 1999, *281* (5), 462–465.

Rice, Thomas H., "The impact of changing Medicare reimbursement rates on physician-induced demand," *Medical care*, 1983, pp. 803–815.

Rodriguez, Juan Carlos, "Measuring financial contagion: A copula approach," *Journal of empirical finance*, 2007, *14* (3), 401–423.

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

Sant'Anna, Pedro HC, "Program evaluation with right-censored data," *arXiv preprint arXiv:1604.02642*, 2016.

Scheike, Thomas H and Mei jie Zhang, "Direct modelling of regression effects for transition probabilities in multistate models," *Scandinavian journal of statistics*, 2007, *34* (1), 17–32.

_ , Mei-Jie Zhang, and Thomas A Gerds, "Predicting cumulative incidence probability by direct binomial regression," *Biometrika*, 2008, *95* (1), 205–220.

Schmid, Christian, "Consumer Health Information and the Demand for Physician Visits," *Health Economics*, 2015, *24* (12), 1619–1631.

Schmid, Christian P. R., Konstantin Beck, and Lukas Kauer, "Health Plan Payment in Switzerland," in Thomas G. McGuire and Richard C. van Kleef, eds., *Risk Adjustment, Risk Sharing and Premium Regulation in Health Insurance Markets*, Academic Press, 2018, chapter 16, pp. 453–489.

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

Schwartz, Aaron L., Bruce E. Landon, Adam G. Elshaug, Michael E. Chernew, and J. Michael McWilliams, "Measuring low-value care in Medicare," *JAMA internal medicine*, 2014, *174* (7), 1067–1076.

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

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

Stano, Miron, "An analysis of the evidence on competition in the physician services markets," *Journal of Health Economics*, 1985, *4* (3), 197 – 211.

Stute, Winfried, "Consistent estimation under random censorship when covariables are present," *Journal of Multivariate Analysis*, 1993, *45* (1), 89–103.

swissinfo, "140 Swiss doctors earn over a million via health insurance," *swissinfo*, 2018. (Available at https://www.swissinfo.ch/eng/tariff-reforms_140-swiss-doctors-earn-over-a-million-via-health-insurance/43886298).

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.

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

Wilkinson, Paul, Sam Pattenden, Ben Armstrong, Astrid Fletcher, R. Sari Kovats, Punam Mangtani, and Anthony J. McMichael, "Vulnerability to winter mortality in elderly people in Britain: population based study," *BMJ: British Medical Journal*, 2004, *329* (7467), 647.

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

Yip, Winnie C., "Physician response to Medicare fee reductions: changes in the volume of coronary artery bypass graft (CABG) surgeries in the Medicare and private sectors," *Journal of Health Economics*, 1998, *17* (6), 675–699.

Zimmer, David M., "The Role of Copulas in the Housing Crisis," *Review of Economics and Statistics*, 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.

Zweifel, Peter, Friedrich Breyer, and Mathias Kifmann, "Health economics," *Berlin/Heidelberg*, 1997, *22009.*

Statement of Authorship

Selbständigkeitserklärung

Ich erkläre hiermit, dass ich diese Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen benutzt habe. Alle Koautorenschaften sowie alle Stellen, die wörtlich oder sinngemäss aus Quellen entnommen wurden, habe ich als solche gekennzeichnet. Mir ist bekannt, dass andernfalls der Senat gemÃ¤ss Artikel 36 Absatz 1 Buchstabe o des Gesetzes vom 5. September 1996 über die Universität zum Entzug des aufgrund dieser Arbeit verliehenen Titels berechtigt ist.

Bern, 15.4.2021

Jonas Christoph Meier