UCLouvain KU LEUVEN

LOUVAIN School of Management

Louvain School of Management
and Katholieke Universiteit Leuven

Nonparametric instrumental regression with right-censored duration outcomes.

A theoretical examination and empirical application on unemployment.

Research Master's Thesis submitted by
Arnaud Germain and Tom Vandorpe

With a view of getting the degrees:
Master Handelsingenieur
Master in Business Engineering, professional Focus

Supervisor :
Ingrid Van Keilegom

Academic Year 2021-2022

Abstract

This thesis studies the effect of cash bonuses on unemployment duration using a non-parametric framework. A discrete instrumental variable independent from the error term of the model and explaining the discrete treatment is used to tackle the endogeneity issue. The proposed model is robust to random right censoring and is applied to the Illinois Unemployment Incentive Experiments data. We estimate quantile treatment effects and show that the efficiency of cash bonuses on unemployment duration varies, depending on the recipient of the bonus (employer or employee) and on the duration that is considered (first spell of unemployment or full benefit year). We also show that the efficiency depends on the ethnicity, gender, age and earnings of an individual by proposing a way to add covariates to the model.

Acknowledgments

We want to thank everyone who helped during the writing process of this thesis. First of all, we would like to express our gratitude to our promotor Prof. Dr. Ingrid Van Keilegom for her guidance and relevant recommendations. Furthermore, we want to thank Dr. Jad Beyhum for answering practical questions about the data and code of his scientific paper. Last but not least, we would like to thank our daily supervisor Lorenzo Tedesco. He helped us a lot in every step of this thesis from coding to writing. He was always available to answer any questions.

Contents

Abstract . . . 1
Acknowledgments . . . 2

1 Introduction **5**

2 Statistical principles used in the proposed model **7**
2.1 Quantile regression . . . 7
2.1.1 Usefulness . . . 7
2.1.2 Principles . . . 8
2.2 Parametric, semi-parametric and non-parametric regression . . . 10
2.2.1 Parametric regression . . . 10
2.2.2 Non-parametric regression . . . 10
2.2.3 Kernel regression . . . 11
2.2.4 Example kernel regression . . . 11
2.3 Instrumental variables . . . 12
2.3.1 Endogenous explanatory variable . . . 12
2.3.2 Requirements for using instrumental variables . . . 12
2.4 Survival analysis . . . 13
2.4.1 Introduction . . . 13
2.4.2 Key concepts . . . 14
2.4.3 Kaplan-Meier estimator . . . 15

3 Theoretical overview of the model **16**

4 Empirical application **20**
4.1 The dataset . . . 20
4.1.1 Dataset description . . . 20
4.1.2 Findings of previous studies . . . 23
4.2 Results . . . 24
4.2.1 Ethnicity . . . 28
4.2.2 Gender . . . 30
4.2.3 Age . . . 32
4.2.4 Earnings before, during and after unemployment . . . 33

5 Conclusion **38**
5.1 Summary of the results . . . 38
5.2 Limits and future research . . . 38

A Figures **42**

B R code **47**

B.1 The model and descriptive statistics 47

B.2 Bootstrapping procedure 53

B.3 Numerical example kernel regression 57

Bibliography **59**

Chapter 1

Introduction

The aim of this thesis is to study the Illinois Unemployment Incentive Experiments by choosing an appropriate model and motivating the choice. Unemployment is extensively studied in the economic literature, because it is an issue of major importance for policymakers. Card and Krueger (2000) (note that Card has been awarded the 2021 Nobel Memorial Prize in Economic Sciences) studied the relationship between minimum wage and unemployment and showed that an increase in the minimum wage does not necessarily imply an increase in unemployment. Furthermore, Chetty (2008) studied the influence of unemployment insurance (UI) benefits on job search behaviour. He showed that increases in unemployment durations due to UI benefits are caused by a liquidity effect rather than by distortions on marginal incentives to search.

One of the policies that has been studied (and even implemented for instance in Belgium), is to reduce UI benefits for long-term unemployed in order to reduce unemployment duration. However, the famous paper Kolsrud et al. (2018) provides evidence that this might not be very effective. Compared to those who have been unemployed for more than six months, the effect of financial incentives is at least three times stronger in the first and second months and twice as strong in the third month of unemployment. The authors therefore conclude that reducing benefits for the short-term unemployed is more effective than reducing benefits for the long-term unemployed.

Instead of "punishing" long-term unemployed, a similar idea consists in "rewarding" an individual who finds a job (or "rewarding" the person who gives the job). Four cash bonus experiments were conducted in the USA: Illinois UI Incentive Experiments (the one we will investigate), New Jersey UI Reemployment Demonstration, Pennsylvania Reemployment Bonus Demonstration and Washington Reemployment Bonus Experiments. Those data sets have been studied several times by researchers to quantify the efficiency of those kinds of incentive programs (an overview can be found in Section 4.1.2). In general, previous findings do not distinguish results in function of the socio-demographic characteristics of the individuals. However, policymakers could be interested in the differences in efficiency between subgroups of the whole population. For instance, is the effect of the treatment the same for men and women? Does the ethnicity of a person influence results? Is the effect the same depending on the age of the participant? How do the earnings of an individual influence the results?

Different methods have been proposed in the survival analysis literature to estimate duration and each has its advantages and drawbacks. Wei et al. (2021) study the complier quantile causal effect (CQCE) using instrumental variable framework with randomly censored outcomes. Besides providing useful insight into the dynamics of the causal

treatment effect, one of the main advantages of the CQCE is that it is identifiable under weaker conditions than the complier average causal effect in the context of censoring. Robust non-parametric estimation is proposed to find weights that can be incorporated into the standard censored quantile regression procedure to estimate CQCE. The estimator exhibits good asymptotic properties and satisfactory finite-sample performance.

Chernozhukov et al. (2011) developed a semi-parametric framework to estimate consistently the quantile regression function of the whole population. The model includes a control variable approach to incorporate endogenous regressors. The censoring duration is assumed to be observed in their framework, but in practice it may be unobserved for some uncensored observations when censoring is caused due to unforeseeable events. The treatment is also assumed to be continuous, whereas our data set contains a discrete treatment.

Sant'Anna (2016) provides a unified framework to estimate average, distributional, and quantile effects on duration in the context of right censoring. The proposed estimators have several advantages: there is a closed-form representation, the implementation is easy and it is fully data-driven upon estimation of nuisance parameters. Moreover, it does not rely on parametric distributional assumptions, shape restrictions or on restricting the potential treatment effect heterogeneity across different subgroups. The method is based on a monotonicity assumption, stating that there are no defiers (subjects who will do the opposite of their treatment assignment status). However, this condition has been criticized in the literature because it restricts the treatment allocation mechanism (Chaisemartin, 2017). Another drawback of this monotonicity assumption is that it enables to estimate local average treatment effects on the population of compliers, but it does not enable to estimate directly the quantile regression function of the whole population such as the rank invariance assumption used in Beyhum et al. (2021).

Lastly, another possible estimator is the one proposed by Beyhum et al. (2021), which analyzes the effect of a discrete treatment (Z) on a duration (T) using quantile treatment effects (QTEs). The model makes use of a discrete instrumental variable (W) in order to resolve the confounding issue, which can be created by the fact that participants can choose to participate or not in the treatment. Furthermore, the model is non-parametric and so no a priori structure is imposed on the model. In addition, the proposed estimation method is robust under random right-censoring, which complies with the structure of the unemployment duration data analyzed.

We use the framework proposed by Beyhum et al. (2021) because it is ideally suited for our application, as we have demonstrated. We propose to enhance this model in order to answer questions raised previously about differences in efficiency of the treatments between subgroups of the whole population, by adding covariates to the model. The covariates can be added to the model in a very intuitive way, which is demonstrated in Chapter 3. Note that the estimator of Wei et al. (2021) is also suited in our context. Therefore it would be interesting to enhance this framework as well in order to include covariates and compare the results with the results we obtain in this thesis.

The thesis is organized as follows. In Chapter 2, the main statistical principles used in the model are discussed briefly. Next, a theoretical overview of the model is provided in Chapter 3. The last part of the thesis consists of the empirical application on unemployment data, reported in Chapter 4, followed by a conclusion in Chapter 5.

Chapter 2

Statistical principles used in the proposed model

In this chapter, a brief explanation of the main statistical principles that are used in the proposed model (see Chapter 3) will be given. The discussed topics are: quantile regression, differences among parametric, semi-parametric and non-parametric regression and the concept of instrumental variables. Lastly the scope of survival analysis will be introduced.

2.1 Quantile regression

2.1.1 Usefulness

Regression techniques are widely used to research a mathematical relation between the measurement of a dependent variable Y and one or more independent variables X, such that the value of the variable Y can be estimated by a measurement of the other variables X. The dominant regression technique is ordinary least squares regression (OLS), which estimates the coefficients of a linear regression model by minimizing the sum of squared errors (SSE). Assume we have n observations $\{(x_i, y_i)\}_{i=1}^n$ for the observable (X, Y), the errors are defined as the differences between the observed outcomes y_i and the outcomes $x_i^T\beta$ predicted by the model. We obtain:

$$SSE = \sum_{i=1}^{n}(y_i - x_i^T\beta)^2. \tag{2.1}$$

The aim of OLS is to calculate the conditional mean of the dependent variable and it is assumed that the relation between the covariates and the dependent variable is the same for the entire distribution. This means that the coefficients are fixed at all level, which might lead to biased representation for some parts of the distribution (Le Cook and Manning, 2013). For example, Coate and Grossman (1988) have shown that the effect of an increase of the alcohol tax significantly reduces the general alcohol consumption level. This decrease in alcohol usage was mainly due to the decreased alcohol consumption of moderate drinkers, while heavy drinkers and light drinkers did only have a very small impact on this decrease. These different impacts on subgroups are not noticed when OLS is used. The resulting model assumes that an alcohol tax has a significant effect on alcohol use for all users, while this might not be true for light and heavy drinkers. Therefore, OLS should be adjusted in order to account for the heterogeneity effects.

For this purpose, the so called quantile regression technique is considered. This estimation method is able to show the different effects on different subsections of the sample, providing a more complete picture of the relationship between the dependent and the independent variables. Moreover, quantile regression also has further advantages. For instance, it is less sensitive to outliers and can deal better with heteroskedasticity. Lastly, it can be used to study the tails of the outcome distribution.

2.1.2 Principles

The previous section stressed the importance and usefulness of quantile regression and now the main principles of this estimation method will be discussed.

It is firstly needed to define the notion of quantiles. The τ^{th}-quantile, where $\tau \in [0, 1]$, is a point that cuts a distribution in 2 parts with proportions τ and $1 - \tau$. For instance, assuming that there is a sample with 10 different observations, the 90^{th} percentile splits the sample in a group with the 9 lowest values and a group with only the highest value. This is illustrated in Figure (2.1) below.

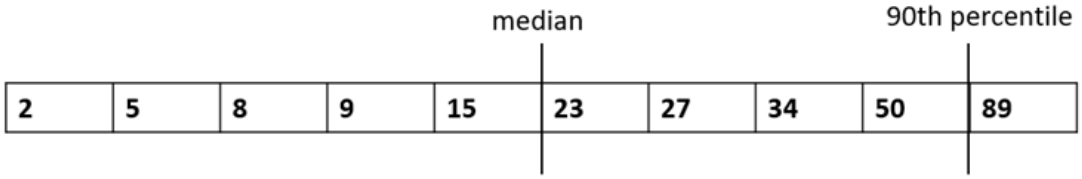

Figure 2.1: Ordered sample divided in quantiles

It is clear that the τ^{th}-quantile is the value y which satisfies:

$$F(y) = \mathbb{P}(Y \leq y) = \tau, \tag{2.2}$$

where Y is a random variable with cumulative distribution $F(y)$. The quantile function $Q(\tau)$ is defined as the inverse of the cumulative distribution function if it exists. Therefore

$$Q(\tau) = F^{-1}(\tau). \tag{2.3}$$

From Equations (2.2) and (2.3) it can be deduced that:

$$\mathbb{P}(Y \leq Q(\tau)) = F(Q(\tau)) = F(F^{-1}(\tau)) = \tau. \tag{2.4}$$

This means that there is a chance of $\tau \times 100\%$ that Y will be below $Q(\tau)$ and a chance of $(1 - \tau) \times 100\%$ that Y will be above $Q(\tau)$. Observe that Equation (2.3) can only be used when F is invertible and continuous. For a general distribution function $F(\cdot)$ we define $Q(\tau)$ as:

$$Q(\tau) = \inf\{y : F(y) \geq \tau\}. \tag{2.5}$$

The basic idea of quantile regression is to estimate the quantile function for a given τ. For instance, for a median regression ($\tau = 0.5$), we seek for the regression line for which 50% of the data points lie below and 50% of the data points lie above the regression line. Observe that this is equivalent to minimizing the sum of absolute errors (SAE), defined as (Davino et al., 2014, p.2):

$$SAE = \sum_{i=1}^{n} |y_i - \hat{y}|, \tag{2.6}$$

where y_i is the actual observation and $\hat{y}$ is the value estimated by the quantile estimator. The extension to a quantile different from the median value is possible using the appropriate loss function $\rho_\tau(u)$ (Koenker, 2005, Chapter 1):

$$\rho_\tau(u) = u(\tau - I(u < 0)) \tag{2.7}$$

where $I(\cdot)$ is the indicator function and u is the error. The aim of the loss function is to weight the errors asymmetrically (see Figure (2.2)), depending on the quantile and the sign of the errors (Dye, 2020). The negative errors are assigned a weight of $(\tau - 1)$ while the positive errors are assigned a weight of τ.

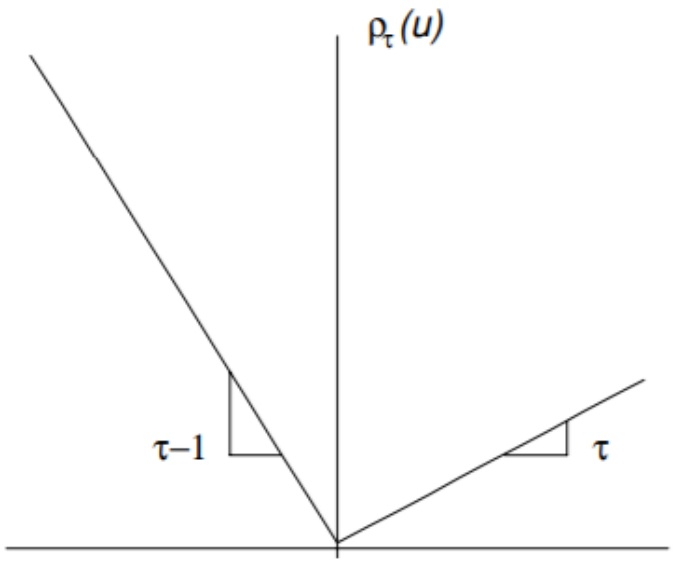

Figure 2.2: Asymmetric loss function $\rho_\tau(u)$ (Koenker, 2005, Chapter 1)

Then, it is possible to find the optimal $\hat{y}$ by minimizing Equation (2.8) below (Koenker, 2005, Chapter 1).

$$E[\rho_\tau(Y - \hat{y})] = (\tau - 1)\int_{-\infty}^{\hat{y}} (y - \hat{y})\, dF(y) + \tau \int_{\hat{y}}^{\infty} (y - \hat{y})\, dF(y) \tag{2.8}$$

Differentiating with respect to $\hat{y}$, results in:

$$0 = (1 - \tau)\int_{-\infty}^{\hat{y}} dF(y) - \tau \int_{\hat{y}}^{\infty} dF(y) = F(\hat{y}) - \tau. \tag{2.9}$$

Equation (2.9) demonstrates that $\hat{y}$ is optimal when $F(\hat{y}) = \tau$, which was already indicated by Equation (2.2) earlier. When the solution for $\hat{y}$ is unique, $\hat{y} = F^{-1}(\tau)$. Otherwise, the smallest value from an interval of τ quantiles must be chosen. According to the loss function $\rho_\tau(u)$, the cost of an underestimation is equal to τ, while the absolute cost of an overestimation is equal to $|\tau - 1| = 1 - \tau$. For instance, when an underestimation of y is three times more costly than an overestimation: we have $\tau = 3(1 - \tau)$ and therefore $\tau = 0.75$ (Koenker, 2005, Chapter 1).

2.2 Parametric, semi-parametric and non-parametric regression

The most well-known category of regression analysis is parametric regression, but when the sample size is large enough, semi-parametric and non-parametric regression are good alternatives with both advantages and disadvantages. In this section these three estimation categories will be discussed and a small example of a non-parametric regression technique will be given.

2.2.1 Parametric regression

When performing regression, the relationship between the explanatory variables (X) and dependent variable (Y) can take on an infinite number of forms. In parametric regression, the functional form of the model is completely assumed by the researcher and only the parameters ($\beta's$) need to be estimated, which means that all other conceivable relationship forms between X and Y are ruled out. One of the most simple forms is a linear regression model, for instance:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon. \tag{2.10}$$

In Equation (2.10), it is clear that the functional form of the model is completely specified, which makes that only β_0, β_1 and β_2 need to be estimated. To solve a linear model, OLS or quantile regression can be used.

A parametric model can also be nonlinear in the parameters, for example:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2^{\beta_3} + \epsilon. \tag{2.11}$$

The parameters of the nonlinear model can not be estimated using OLS. Instead, another estimation method (e.g. nonlinear least squares estimation) needs to be used to solve the nonlinear model (Mahmoud, 2019). Parametric models are very intuitive and they are easy to estimate and interpret. The downside of this regression method is that it only generates reliable results when the functional form is specified correctly. The researcher can try to guess the correct form of the model by looking at the data, but when the assumed form does not fit the data well, the estimates will not be reliable.

2.2.2 Non-parametric regression

In non-parametric regression, the regression function is estimated without specifying the functional form of the model beforehand, which means that no functional form is ruled out a priori. Let us define the non-parametric regression model as $Y = f(X) + \epsilon$, where $f(.)$ is the unspecified regression function. There are no assumptions made about $f(.)$, except that it needs to be a smooth function. This ensures that non-parametric regression estimators are very flexible. However, a drawback of this estimation method is that the precision of a non-parametric regression estimator decreases drastically when multiple covariates are added to the model. Such problems are usually called "curse of dimensionality". To deal with it, researchers developed another estimation method called the semi-parametric model. This method combines features of the parametric model and the non-parametric model to obtain a technique that can handle many covariates and is as flexible as possible.

2.2.3 Kernel regression

The two most common non-parametric estimation techniques are kernel regression and spline smoothing (Mahmoud, 2019). Only kernel regression will be discussed in this dissertation because this technique is used in the empirical application. The most popular method for kernel regression is the Nadaraya-Watson estimator, which tries to estimate the unknown function $f(x)$ in non-parametric regression (Mahmoud, 2019). The Nadaraya-Watson estimator gives a weight (w_i) to all the points in the data set based on the distance between x_i and x, where x_i is the observed value and x is the value for which $f(x)$ is computed. Data points for which the distance between x_i and x is small will get a high weight and vice versa. The Nadaraya-Watson estimator is given in Equation (2.12) below.

$$\widehat{f}_h(x) = \frac{\sum_{i=1}^{n} K\left(\frac{x-x_i}{h}\right) y_i}{\sum_{i=1}^{n} K\left(\frac{x-x_i}{h}\right)} = \sum_{i=1}^{n} w_i y_i \tag{2.12}$$

where $K(u)$ is the kernel function with $u = \frac{x-x_i}{h}$ and h is the bandwidth. The kernel function $K(u)$ must be symmetrical $(K(u) = K(-u))$ and the area under the curve must be equal to one $(\int_{-\infty}^{+\infty} K(u)du = 1)$ (Pramanik, 2019). Another property of $K(u)$ is that this function can not be negative, so $K(u) \geq 0$ for all values of u. There exist several different types of kernel functions like the Gaussian kernel or the Epanechnikov kernel. In this section, the Epanechnikov kernel will be used to illustrate how kernel regression works, since this is also the kernel used in the proposed model in Chapter 3. The Epanechnikov kernel is defined as:

$$K(u) = \frac{3}{4}(1-u^2) \quad with\ |u| \leq 1. \tag{2.13}$$

The bandwidth (h) regulates the smoothness of $\hat{f}(x)$. When h is too small, $\hat{f}(x)$ will be over-fitted, while $\hat{f}(x)$ will be over-smoothed when h is too large. In the latter case the true relationship between $f(x)$ and x will not be revealed (Pramanik, 2019).

2.2.4 Example kernel regression

To illustrate the theory discussed above, a simple example will be given. Assume that y represents the number of citizens of a city (in thousands), while x is representing the area of the city (in km^2). Table (2.1) below shows a fictional data set with six observations.

obs.	y	x
1	212	230
2	184	120
3	168	100
4	45	45
5	30	35
6	10	20

Table 2.1: Numerical example for kernel regression

Assume that $x = 60$ and $\hat{f}(x)$ is calculated based on the values in Table (2.1) using Epanechnikov kernel regression. The optimal bandwidth is calculated using the h.amise function in R and is approximately equal to 315. Subsequently the weights w can be calculated, using the Epanechnikov kernel function $K(u)$.

$$w_i = \frac{K\left(\frac{60-x_i}{315}\right)}{\sum_{i=1}^{6} K\left(\frac{60-x_i}{315}\right)} \tag{2.14}$$

The values of the weights for $x = 60$ are given in Table (2.2) below and the sum of the weights is equal to one.

w_1	w_2	w_3	w_4	w_5	w_6
0.1258500	0.1711256	0.1747047	0.1771654	0.1764495	0.1747047

Table 2.2: Weights of the kernel regression

Once the weights are calculated, it is very easy to estimate $f(60)$ using Equation (2.12).

$$\hat{f}(60) = 0.1258500 \cdot 212 \ + \ 0.1711256 \cdot 184 \ + \ 0.1747047 \cdot 168 \ + \ 0.1771654 \cdot 45 \ + \ 0.1764495 \cdot 30 \ + \ 0.1747047 \cdot 10 = 102.53 \tag{2.15}$$

Additionally, $\hat{f}(x)$ can be calculated for all values of x by following the steps described in this section.

2.3 Instrumental variables

In a regression model, an explanatory variable is referred to as endogenous if it is correlated with the error term, which can lead to biased estimates of the explanatory variable. The most common way to resolve the issue of endogeneity is to use instrumental variables. This technique and the various causes of endogeneity will be discussed in this section.

2.3.1 Endogenous explanatory variable

The three main reasons for endogeneity are: measurement errors, simultaneity and omitted confounding variables. Only confounding variables will be discussed in this dissertation, as this is the issue which occurs in the empirical application. A confounding variable (U) is a variable, which has an effect on the explanatory variable (X) and on the dependent variable (Y), and therefore obscures the real effect of X on Y.
To clarify this, a small example will be given about a study that compares the effectiveness of a treatment for premature babies in high technology neonatal intensive care units (high level NICU's) with the effect of a treatment in local hospitals (low level NICU's). As stated in Baiocchi et al. (2014), the baby's health before arrival in the hospital is a good example of a confounder, because it is correlated with the explanatory variable (babies with more severe health problems are more likely to be treated in a high level NICU) and with the outcome variable (the treatment of babies with more severe heath problems is more likely to be less effective). As a consequence, there is a difference between the treatment groups before the treatment has started and this difference does affect the outcome. To solve this problem, an instrumental variable can be used to obtain consistent estimates of the treatment variable.

2.3.2 Requirements for using instrumental variables

An instrumental variable (W) needs to fulfill the following three requirements:

- W is independent of U.
- W has a causal effect on X.
- W only affects Y indirectly through its effect on X.

In Figure (2.3) the causal relationships between W, X, U and Y are given and all the requirements of the instrumental variable (W) are fulfilled.

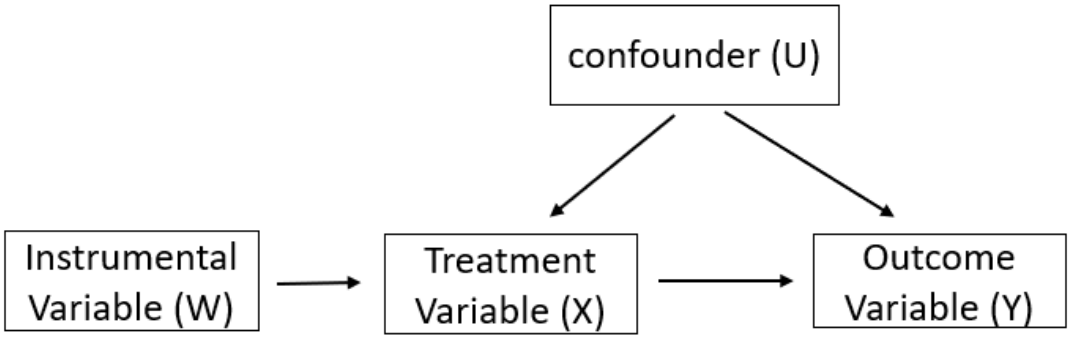

Figure 2.3: Causal relationship between W, X, U and Y

In Baiocchi et al. (2014) the excess travel time, which is defined as the time it takes to travel to the nearest high level NICU minus the time it takes to travel to the nearest low level NICU, is considered as a possible instrumental variable in order to resolve the problem of the unmeasured confounder. It is proposed that $W = 1$ if the excess travel time is smaller than 10 minutes and $W = 0$ otherwise. The excess travel time (W) will have a causal effect on the choice of the treatment (X), because the chance that a baby will be delivered in a high level NICU will be higher when $W = 1$ than when $W = 0$. In addition, the excess travel time is independent of the baby's health before arrival in the hospital and does only influence Y through its effect on the choice of the treatment. Thus, the excess travel time does indeed meet all the requirements for an instrumental variable, which are stated above.

2.4 Survival analysis

2.4.1 Introduction

In survival analysis, the outcome variable (T) represents the time until the occurrence of a certain event of interest, also called the survival time or spell length (Clark et al., 2003). This event can be anything, for example in our empirical analysis (Chapter 4), the survival time is the time it takes before a participant finds a job. It is possible that a participant does not find a job before the study is ended or drops out of the study due to another reason than finding a job, which makes it impossible to know exactly how long the participant remained unemployed. In the former case it is only known that T is larger than the duration of the study while in the latter case it is only know that T is larger than the moment the participant dropped-out (assuming that the start of the spell is equal to the start of the study). In these two cases the observation is called right-censored (Jenkins, 2005). In addition, it can happen that the starting time of a spell is unknown, which makes the observation a left-censored observation. Furthermore, survival data are often skewed and consist typically of many early events and relatively few late events. Traditional methods such as OLS are not well suited to deal with censoring and data with a skewed distribution. Therefore, survival analysis is characterized by the usage of different techniques than OLS, as will be explained below.

2.4.2 Key concepts

To demonstrate the modeling of survival analysis, it is needed to introduce: the survival function, the failure function and the hazard rate. The survival time (T) has a cumulative distribution $F(t)$, also called the failure function (Jenkins, 2005):

$$F(t) = \mathbb{P}(T \leq t). \tag{2.16}$$

The survival function $S(t)$ indicates the probability that the event of interest did not yet happen at time t and has a one-to-one relationship with the failure function. The survival function has values between 0 and 1 and is strictly decreasing, starting from one:

$$S(t) = \mathbb{P}(T > t) = 1 - F(t). \tag{2.17}$$

For instance, in our empirical application, if $S(12) = 0.45$, then there is a chance of 45% that an individual did not find a job after 12 weeks.

To model the hazard rate it is needed to introduce the probability density function $f(t)$, which is the probability that an event occurs between time t and $t + \Delta t$, where Δt is infinitesimally small (Jenkins, 2005):

$$f(t) = \lim_{\Delta t \to 0} \frac{\mathbb{P}(t \leq T \leq t + \Delta t)}{\Delta t} = \frac{\partial F(t)}{\partial t}. \tag{2.18}$$

The hazard rate $h(t)$ is constructed using $f(t)$ and $F(t)$ and represents the chance that the event occurs between t and $t + \Delta t$, conditional on survival up to time t:

$$h(t) = \frac{f(t)}{1 - F(t)} = \frac{f(t)}{S(t)}. \tag{2.19}$$

The concepts above will be clarified with a simple example. Assume that there are six participants in a study with a duration of five days. The event times (in days) are: {1, 3, 4, 4, 5, 8}. Table (2.3) below shows how $F(t)$, $S(t)$, $f(t)$ and $h(t)$ evolve over time.

t	F(t)	S(t)	f(t)	h(t)
0	0	1	0	0
1	1/6	5/6	1/6	1/6
2	1/6	5/6	0	0
3	2/6	4/6	1/6	1/5
4	4/6	2/6	2/6	1/2
5	5/6	1/6	1/6	1/2

Table 2.3: Numerical example for survival analysis

2.4.3 Kaplan-Meier estimator

The survival function $S(t)$ can be estimated by means of the non-parametric Kaplan-Meier estimator (Kaplan and Meier, 1958), which does not make any parametric assumptions about the theoretical distribution of $S(t)$ a priori. Assume that $t_1 < t_2 < ... < t_n$ are the survival times of the data set and the number of subjects at risk just before time t_i are represented by n_i. Thus, n_i are all individuals that are not censored and for whom the event of interest did not yet occur just before time t_i. The number of events that happen exactly at time t_i are represented by d_i. The Kaplan-Meier estimator is:

$$\widehat{S}(t) = \prod_{i|t_i \leq t} \left(1 - \frac{d_i}{n_i}\right). \tag{2.20}$$

For example $\widehat{S}(t_1)$ is simply equal to $1 - \frac{d_1}{n_1}$ where d_1 is the number of events that happened at t_1 and n_1 is the number of subjects at risk just before t_1 (Jenkins, 2005, Chapter 4). In addition $\widehat{S}(t_2)$ is calculated by multiplying $\widehat{S}(t_1)$ with $1 - \frac{d_2}{n_2}$ and all other survival states can be calculated in the same manner. To illustrate this the numerical example from Table (2.3) can be used. Calculating $\widehat{S}(t_4)$ can be done simply by multiplying $\widehat{S}(t_3)$ with $1 - \frac{d_4}{n_4}$, which is equal to $\frac{4}{6} \cdot (1 - \frac{2}{4}) = \frac{2}{6} = S(t_4)$.

Chapter 3

Theoretical overview of the model

The theoretical model is based on the one provided by Beyhum et al. (2021). We enhance their framework to include covariates X. Let us define:

$$T = \varphi(Z, X, U) = \varphi_{Z,X}(U), \tag{3.1}$$

where $\varphi : \{z_1, ..., z_L\} \times \{x_1, ..., x_P\} \times \mathbb{R}_+ \to \mathbb{R}_+$ is a quantile regression function that is square integrable with respect to the distribution of $(Z,\ X,\ U)$, T is the duration variable, U is a random element with unit exponential distribution, Z is a categorical treatment variable with support $\{z_1, ..., z_L\}$ and X is a random covariate with support $\{x_1, ..., x_P\}$. Furthermore, the variable U represents both ex-ante (prior to the duration spell) differences among individuals and ex-post shocks occurring during duration. For each z, x and for all positive real values of u, $\varphi_{z,x}(u)$ is the $1 - e^{-u}$ quantile of the duration for treatment z and covariate value x. From φ the Quantile Treatment Effect (QTE) of a change in treatment from z_0 to z_1 can be derived, for all values of x and all positive values of u:

$$QTE_x(u) = \varphi(z_1, x, u) - \varphi(z_0, x, u). \tag{3.2}$$

Therefore, Equation (3.2) represents the QTE at the $1 - e^{-u}$ quantile. The aim is to estimate the QTE and not only the Average Treatment Effect (ATE) for several reasons. In the case of cash bonuses to reduce unemployment duration (see empirical application in Chapter 4), policymakers could be interested not only in the mean net impact but also in the distributional impacts of the program. For instance, treatments could be significant for short-term unemployment and not for long-term unemployment, or vice-versa. Moreover, not only QTEs but also survival probabilities and hazard rates can be derived from the quantile regression function. Regarding the conditional survival function S, it holds that:

$$S(t \mid z, x) = \mathbb{P}(T \geq t \mid z, x) = \mathbb{P}(\varphi(z, x, U) \geq t). \tag{3.3}$$

Note that the conditional cumulative distribution function F can be defined as a complement of S:

$$F(t \mid z, x) = \mathbb{P}(T < t \mid z, x) = 1 - S(t \mid z, x). \tag{3.4}$$

The hazard function h, which is discussed in Section 2.4, is defined as:

$$h(t \mid z, x) = \frac{F'(t \mid z, x)}{S(t \mid z, x)} = \frac{\frac{\partial \mathbb{P}(\varphi(z,x,U) \leq t)}{\partial t}}{\mathbb{P}(\varphi(z, x, U) \geq t)}. \tag{3.5}$$

Note that the ATE can still be derived from the quantile regression function:

$$ATE_x = \mathbb{E}[\varphi(z_1, x, U) - \varphi(z_0, x, U)]. \tag{3.6}$$

The duration is right censored by a random variable C with support in $\mathbb{R}^+$ and we assume that C is conditionally independent of T given $(Z,\ X,\ W)$. We use rank invariance assumption by assuming that for $z \in \{z_1, ..., z_L\}$ and $x \in \{x_1, ..., x_P\}$, the mapping $\varphi_{z,x}(.)$ is strictly increasing and differentiable. This assumption means that the rank in outcomes between two individuals does not change with treatment values. We also define δ, which indicates if a duration is censored or not:

$$\delta = I(T < C). \tag{3.7}$$

We now use the estimator proposed by Beyhum et al. (2021) to estimate the quantile regression function. First, a categorical instrumental variable W with support $\{w_1, ..., w_K\}$ is introduced. By applying Equation (3.3), we get:

$$\begin{aligned}\sum_{l=1}^{L} S(\varphi_{z_l,x_p}(u), z_l \mid x_p, w_k) &= \sum_{l=1}^{L} \mathbb{P}(T \geq \varphi_{z_l,x_p}(u), Z = z_l \mid X = x_p, W = w_k) \\ &= \sum_{l=1}^{L} \mathbb{P}(\varphi_{z_l,x_p}(U) \geq \varphi_{z_l,x_p}(u), Z = z_l \mid X = x_p, W = w_k).\end{aligned} \tag{3.8}$$

Then, the right-hand side of Equation (3.8) can be simplified since, as stated before, the mapping $\varphi_{z,x}(.)$ is strictly increasing:

$$\sum_{l=1}^{L} S(\varphi_{z_l,x_p}(u), z_l \mid x_p, w_k) = \sum_{l=1}^{L} \mathbb{P}(U \geq u, Z = z_l \mid X = x_p, W = w_k). \tag{3.9}$$

If we apply the law of total probability on Equation (3.9), we have:

$$\sum_{l=1}^{L} S(\varphi_{z_l,x_p}(u), z_l \mid x_p, w_k) = \mathbb{P}(U \geq u \mid X = x_p, W = w_k). \tag{3.10}$$

Assuming that W and X are both independent from U, we can rewrite:

$$\sum_{l=1}^{L} S(\varphi_{z_l,x_p}(u), z_l \mid x_p, w_k) = 1 - (1 - e^{-u}) = e^{-u}. \tag{3.11}$$

In addition, if Z is continuous this equation is transformed to:

$$\int S(\varphi_{z,x_p}(u), z \mid x_p, w_k) dz = e^{-u}. \tag{3.12}$$

Let us define the vector function A composed of K elements (where K is the number of levels of the instrumental variable W) as:

$$A(\tilde{\varphi}, \tilde{S})(u) = \Big(\sum_{l=1}^{L} \tilde{S}(\tilde{\varphi}_{z_l,x_p}(u), z_l \mid x_p, w_k) - e^{-u} \Big)_{k=1}^{K}. \tag{3.13}$$

The function φ belongs to the set of solutions of the equations:

$$A(\varphi, S) = 0. \tag{3.14}$$

The estimator $\hat{\varphi}$ is defined as

$$\hat{\varphi} \in \arg\min_{\theta \in F_{Z,X}^{\bar{U},\bar{T}}} \|A(\theta, \hat{S})\|_V^2, \tag{3.15}$$

where $\bar{U} < \infty$ is an upper bound up to which φ is estimated, $\bar{T} < \infty$ is an upper bound on $\max_{l=1,...,L;p=1,...,P} \ \varphi(z_l, x_p, \bar{U})$ and $F_{Z,X}^{\bar{U},\bar{T}} = \{f : \{z_1, ..., z_L\} \times \{x_1, ..., x_P\} \times [0, \bar{U}] \mapsto [0, \bar{T}]\}$. Furthermore, let V(u) be a positive definite $K \times K$ weighting matrix for each $u \in [0, \bar{U}]$. Let us define $\|v\|_2 = \sqrt{v^T v}$ and $\|v\|_{\bar{V}} = \sqrt{v^T \bar{V} v}$ for a vector $v \in \mathbb{R}^K$ and a $K \times K$ matrix $\bar{V}$. It is assumed that the system (3.10) has a unique solution in $F_{Z,X}^{\bar{U},\bar{T}}$ and that $\hat{S}$ is a consistent estimator of S on $[0, \bar{T}]$. Under some assumptions, it can be shown (Beyhum et al., 2021) that this estimator has good asymptotic properties. In particular, it is shown that $\hat{\varphi}$ is asymptotically normal and that the rate of convergence of $\hat{\varphi}$ is equal to the one of $\hat{S}$ considering a uniform norm. The uniform norm assigns to real-valued bounded functions f defined on a set S the non-negative number:

$$\|f\|_\infty = \sup \{ |f(s)| : s \in S \}. \tag{3.16}$$

Next, an estimator for the survival function that is robust to random right censoring is needed. The duration is right-censored at C and $Y = \min(T, C)$. The Kaplan-Meier estimator (Kaplan and Meier, 1958) is consistent to estimate conditional survival functions in the context of random right censoring (Dabrowska, 1989) and is defined as:

$$\hat{S}_{KM}(t \mid z, x, w) = \prod_{s \leq t} \Big(1 - \frac{dN_{z,x,w}(s)}{Y_{z,x,w}(s)}\Big), \tag{3.17}$$

where

$$Y_{z,x,w}(t) = \sum_{i=1}^{n} I(Y_i \geq t, Z_i = z, X_i = x, W_i = w), \tag{3.18}$$

and where $dN_{z,x,w}(s)$ denotes the jump of the process

$$N_{z,x,w}(t) = \sum_{i=1}^{n} I(Y_i \leq t, Z_i = z, X_i = x, W_i = w, \delta_i = 1), \tag{3.19}$$

at time s such that:

$$dN_{z,x,w}(s) = N_{z,x,w}(s) - \lim_{s' \mapsto s, s' < s} N_{z,x,w}(s'). \tag{3.20}$$

Next, it is important to smooth the Kaplan-Meier estimator. As shown by Beyhum et al. (2021), smoothing ensures that $\sqrt{n}(\hat{\varphi} - \varphi)$ converges to a mean zero Gaussian process. It is possible to obtain a smooth version of the estimator considering a smoothing kernel approach, as specified in Section 2.2.3. Specifically, we define a kernel K with bandwidth h:

$$\hat{S}_{smooth}(t \mid z, x, w) = \int \hat{S}_{KM}(t - sh \mid z, x, w) K(u) du. \tag{3.21}$$

The conditional smoothed estimator $\hat{S}_{smooth}$ is then transformed into our final estimator:

$$\hat{S}(t, z \mid x, w) = \hat{S}_{smooth}(t \mid z, x, w) \frac{\sum_{i=1}^{n} I(Z_i = z, X_i = x, W_i = w)}{\sum_{i=1}^{n} I(X_i = x, W_i = w)}. \tag{3.22}$$

In practice, optimization program (3.15) is not feasible. Instead φ can be estimated over a grid of M values (with u_m from u_1 to u_M):

$$(\hat{\varphi}(z_l, x_p, u_m))_{l=1}^{L} \in \arg\min_{\theta \in [0, \bar{T}]^L} \|A(\theta, \hat{S})(u_m)\|^2_{V(u_m)}, \tag{3.23}$$

where $\bar{T}$ is the value at which the data set is censored. We propose to use the optimization algorithm of Byrd et al. (2003) because the optimization program for our estimator

is bounded. It is a limited-memory modification of the BFGS quasi-Newton method. In practice, Q realisations of a vector composed of L uniform random variables $U_{[0,\bar{T}]}$ are generated for each value of u, using the Monte-Carlo method. All Q vectors will serve as starting points for the minimization program. For each value of u, the estimated quantiles that lead among all Q values of the objective function to the lowest one are retained.

Pointwise confidence intervals can be build using a bootstrap resampling technique in the following way. First, B resamples of size n are drawn with replacement from the original data set and the value of our estimator φ_b is computed for each resample. Next, a 95% confidence interval for $f((\varphi_{z_l,x_p}(u))_{l=1}^{L})$ is constructed, using the 2.5% and 97.5% percentiles of $\{f((\hat{\varphi}_b(z_l, x_p, u))_{l=1}^{L})\}_{b=1}^{B}$ for the lower and upper bound.

Finally, when the covariate of interest is categorical, it can be added to the model using another very intuitive method. In Beyhum et al. (2021), a very simple technique is proposed to add covariates to the described model by including them in the treatment variable Z. Since the covariate X is now included in Z, Equation (3.11) is simplified to:

$$\sum_{l=1}^{L} S(\varphi_{z_l}(u), z_l \mid w_k) = e^{-u}. \tag{3.24}$$

Only categorical variables can be included in Z, so discrete and continuous covariates need to be transformed to categorical variables before they can be added to the model using this method. In the empirical example in the next chapter, the transformation from discrete and continuous to categorical variables will be demonstrated extensively.

Chapter 4

Empirical application

4.1 The dataset

4.1.1 Dataset description

The data set comes from the Illinois Unemployment Incentive Experiments and can be found on this website: https://www.upjohn.org/data-tools/employment-research-data-center/illinois-unemployment-incentive-experiments. Woodbury and Spiegelman (1987) have provided a detailed description of the experiment. This controlled social experiment was conducted by the Illinois Department of Employment Security in 1984. The goal of the experiment was to test the effectiveness of cash bonuses in reducing the duration of insured unemployment. New claimants for Unemployment Insurance (UI) were randomly assigned to one of these 3 groups:

- The Job Search Incentive Experiment group (JSIE): they would qualify for a cash bonus of 500$ if they find a job of at least 30h/week within 11 weeks from the beginning of their unemployment spell and held that job for 4 months. Claimants have to read the description of the experiment and sign an agreement beforehand to be eligible to the bonus.

- The Hiring Incentive Experiment group (HIE): exactly the same conditions as above but this time the 500$ bonus is given to the company that is hiring the job seeker.

- The control group.

The experiment aimed to give an incentive to job seekers to search more intensely for work (JSIE) or to provide a marginal wage-bill subsidy (HIE) in order to reduce the duration of unemployment. We use two different variables to express the duration of unemployment throughout our analysis. Firstly, the duration of unemployment can be understood as the number of weeks of UI during the first spell of unemployment (immediately following the initial claim). Secondly, it can be expressed as the number of weeks of UI during a full benefit year. Note that for each individual the number of weeks for the full year is always larger or equal to the number of weeks of the first spell. The two durations will be different if and only if an individual who has found a job becomes unemployed again during the year. Furthermore, individuals assigned to the control group did not know about the experiment. Table (4.1) shows that the characteristics of the participants are equally distributed over the treatment groups, due to randomization.

	Control group	JSIE	HIE
Age	32.980	32.928	33.099
Male	0.547	0.563	0.538
Black	0.271	0.251	0.256
White	0.632	0.651	0.647
Hispanic	0.076	0.074	0.077
Native American	0.006	0.010	0.006
Others	0.015	0.015	0.014
Base period earnings	3188.308	3221.980	3214.577

Table 4.1: Mean of different variables for each group

Each individual assigned to a group could refuse (non-complier) or agree (complier) to participate. The goal of this thesis is to analyze the effect of the 2 different treatments on the duration of unemployment. Let us define the treatment variable Z:

$$Z = \begin{cases} 2 \text{ if the person is assigned to HIE group and agrees to participate} \\ 1 \text{ if the person is assigned to JSIE group and agrees to participate} \\ 0 \text{ if the person is assigned to control group or refuses to participate.} \end{cases}$$

We want to analyze the effect of the discrete treatment Z on the unemployment duration, which we denote by T. The average unemployment duration $\overline{T}$ of each group is reported in Tables (4.2) and (4.3). The mean of JSIE and HIE group is lower than the one of the control group. Moreover, the mean duration for the compliers is lower than the mean duration for the non-compliers. Note also that the average unemployment duration of the control group and the non-compliers are close to each other. Those preliminary results suggest that both treatments tend to reduce unemployment duration (whatever duration you consider).

	All	Control group	JSIE			HIE		
			All	Complier	Non-complier	All	Complier	Non-complier
$\overline{T}$	17.63	18.33	16.96	16.74	18.18	17.65	17.62	17.72

Table 4.2: Average unemployment duration (first spell) of each group

	All	Control group	JSIE			HIE		
			All	Complier	Non-complier	All	Complier	Non-complier
$\overline{T}$	19.54	20.06	18.91	18.71	19.98	19.70	19.67	19.75

Table 4.3: Average unemployment duration (full year) of each group

A t-test can be used, in order to check if differences between groups are significant. Since groups are independent and sample sizes are unequal, a Welch t-test is used. The test statistic is given by the following formula:

$$t = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{s^2_{\overline{X}_1} + s^2_{\overline{X}_2}}}, \tag{4.1}$$

where $\overline{X}$ and s^2 are the sample mean and the sample variance. The p-value for the Welch t-test between control group (Z=0) and JSIE compliers (Z=1) for the full year is smaller than 0.001 whereas the p-value between control group (Z=0) and HIE compliers (Z=2) is 0.28. Since the null hypothesis assumes equal means between groups,

that means that the difference between JSIE and control group is significant and the difference between HIE and control group is not significant at a 95% confidence level for the full year duration. For the first spell, the p-value between control group and JSIE compliers is smaller than 0.001 whereas the p-value for the HIE compliers is 0.07. The difference for the JSIE treatment is therefore significant but the one for the HIE group is ambiguous because it is significant at a 90% confidence level but not significant at a 95% confidence level.

The decision to not participate in a bonus program is very likely to be correlated to the unobserved motivation to find work, which in turn is correlated with the unemployment duration. This makes that the unobserved motivation is a unmeasured confounder (see section 2.3) and that the treatment variable Z is endogenous (Bijwaard, 2007). The decision to participate is endogenous even if individuals are allocated randomly to the groups. This could induce selection bias that could be due to the extra administration and communication that is necessary to be part of a bonus program. The first problem we have to deal with is thus the presence of endogeneity. Similarly to OLS, endogeneity of explanatory variables in the context of quantile regression leads to inconsistency in the estimation of QTEs (Chernozhukov and Hansen, 2013). According to Baiocchi et al. (2014), the random assignment can be used as an instrumental variable, when the decision to participate is endogenous. Therefore, W is introduced to deal with the problem of endogeneity, using group assignment:

$$W = \begin{cases} 2 \text{ if the person is assigned to HIE group} \\ 1 \text{ if the person is assigned to JSIE group} \\ 0 \text{ if the person is assigned to control group.} \end{cases}$$

About 15% of the participants that were assigned to the JSIE group and 35% of the participants that were assigned to the HIE group refused to participate in the experiment (see the sample sizes of each combination of Z and W in Table (4.4)). The data set is composed of 12,101 observations. Refusal rates suggest large selection bias and underline the usefulness of the instrumental variable.

	W =	0	1	2
	0	3952	659	1377
Z=	1	0	3527	0
	2	0	0	2586

Table 4.4: Sample sizes of the experiment

The duration is expressed as the number of weeks of UI and is therefore discrete. However, for simplicity we will treat the duration variable as continuous in this thesis. Moreover, the duration is censored at 26 weeks, because UI was granted at most for 26 weeks. Around 40% of the observations are censored. Note that the assumption of rank invariance allows that job seekers derive heterogeneous benefits from the treatments, but rules out that an individual a, for which the duration of unemployment without treatment is larger than the one of an individual b, benefits enough from the treatment to find a job before individual b.

4.1.2 Findings of previous studies

Woodbury and Spiegelman (1987) were the first to analyze the data of this experiment. They observe a benefit reduction of $158 to $194 (depending on whether Federal Supplemental Compensation benefits are included) and a 1.15-week reduction in the duration of unemployment for the JSIE group compared to the control group. Those differences are statistically different from 0 at a 1-percent level and they were attained on average over all participants, whether or not they agreed to participate or to obtain the bonus. Furthermore, the number of participants that ended their spell of insured unemployment within 11 weeks of filing, was 5.5% higher in the JSIE group compared to the control group. Those results suggest that the JSIE treatment reduces the duration of unemployment. They report quite different results for the HIE group, where there are no statistical differences with the control group regarding benefit and duration of unemployment. Note that this is only true if the full benefit year is considered, while there is a reduction of the unemployment duration when only the duration of unemployment during the first spell is considered. They also stress the problem that eligible claimants refused to participate and therefore "results could be biased estimates of the experimental effect".

Meyer (1996) investigates the hazard rate (rate at which unemployment ends at some time T given that it has lasted until T) using a Proportional Hazard model with a piecewise constant baseline hazard. Hazard rates of JSIE group are consistently above the ones of control group, which suggests once again that cash bonuses tend to reduce the duration of unemployment. Bijwaard and Ridder (2005) use an instrumental variable estimator in a Mixed Proportional Hazard model context. They argue that estimates used by Meyer (1996) are most likely downward biased due to selective non-compliance. In other words, they find an even larger effect of the JSIE on the duration of unemployment. Bijwaard (2007) uses a Generalized Accelerated Failure Time model that encompasses Mixed Proportional Hazard model and Accelerated Failure Time model and finds similar results. Sant'Anna (2020) analyzes the two treatments separately using different non-parametric tests for treatment effect heterogeneity based on two-step Kaplan-Meier integrals. Their results suggest that giving an unemployment bonus to the job searcher changed the length of the unemployment duration, while giving a bonus to the employer is ineffective in changing the unemployment duration. In Beyhum et al. (2021) a non-parametric framework for instrumental quantile regression with right censored duration outcomes is proposed. Their results are in line with previous studies. For most quantiles, the QTE of the JSIE bonus is significant while the QTE of the HIE bonus is not.

4.2 Results

We estimate QTEs of the two treatments (JSIE and HIE) for each duration (first spell and full year) using the model described in the eponymous section. For this first part of the section, we do not include any covariate. The first step to undertake is the estimation of the survival function using the Kaplan-Meier estimator. Estimated curves for the first spell and the full year durations are respectively displayed in Figures (4.1) and (A.1) (see Appendix A for the latter).

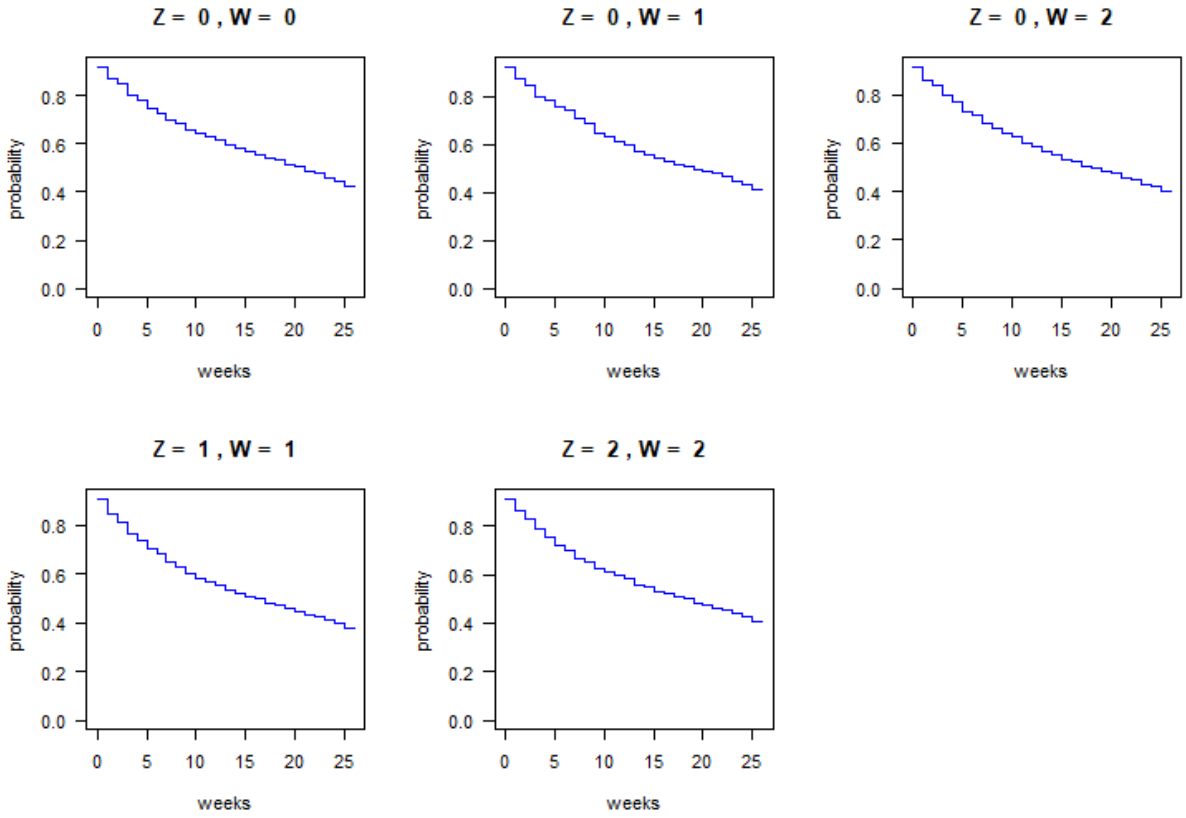

Figure 4.1: Kaplan-Meier estimation for the first spell duration

Once the Kaplan-Meier estimator is obtained, it is smoothed using a kernel smoothing procedure (see Section 2.2.3). We set the bandwidth equal to two and opt to use an Epanechnikov kernel.

$$K(u) = \frac{3}{4}(1 - u^2) \tag{4.2}$$

The final estimator for the survival function can be obtained using Equation (3.22). Estimated survival functions for the first spell and the full year durations are respectively displayed in Figures (4.2) and (4.3).

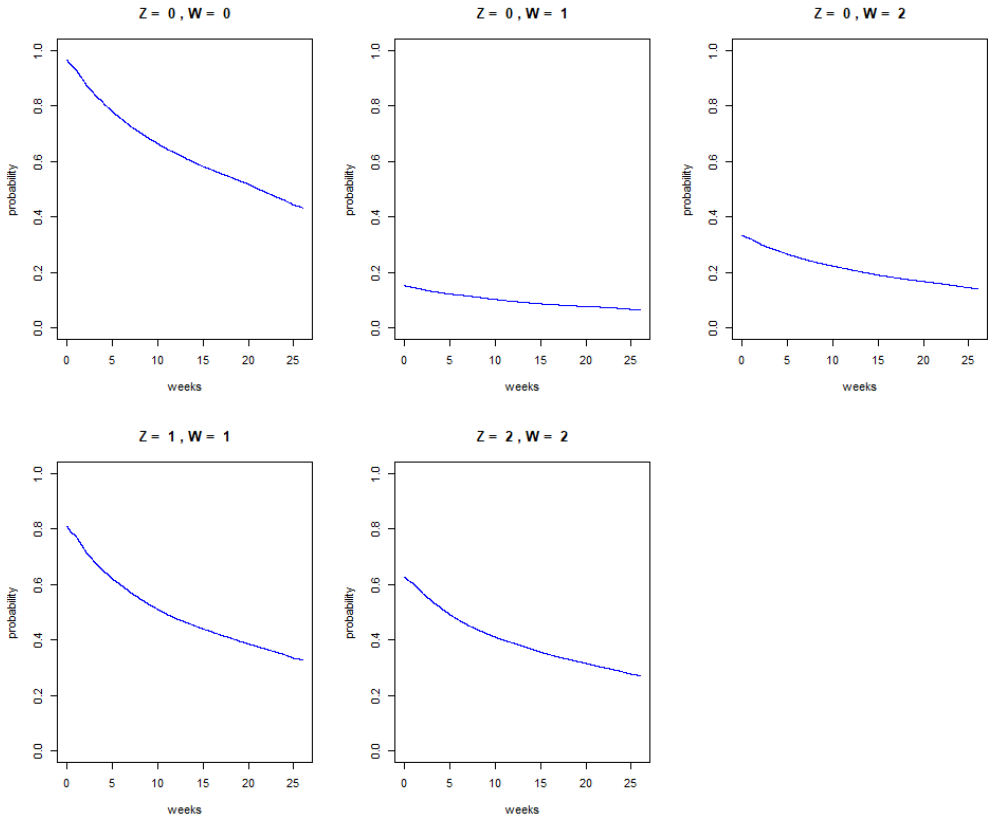

Figure 4.2: Survival function estimation for the first spell duration

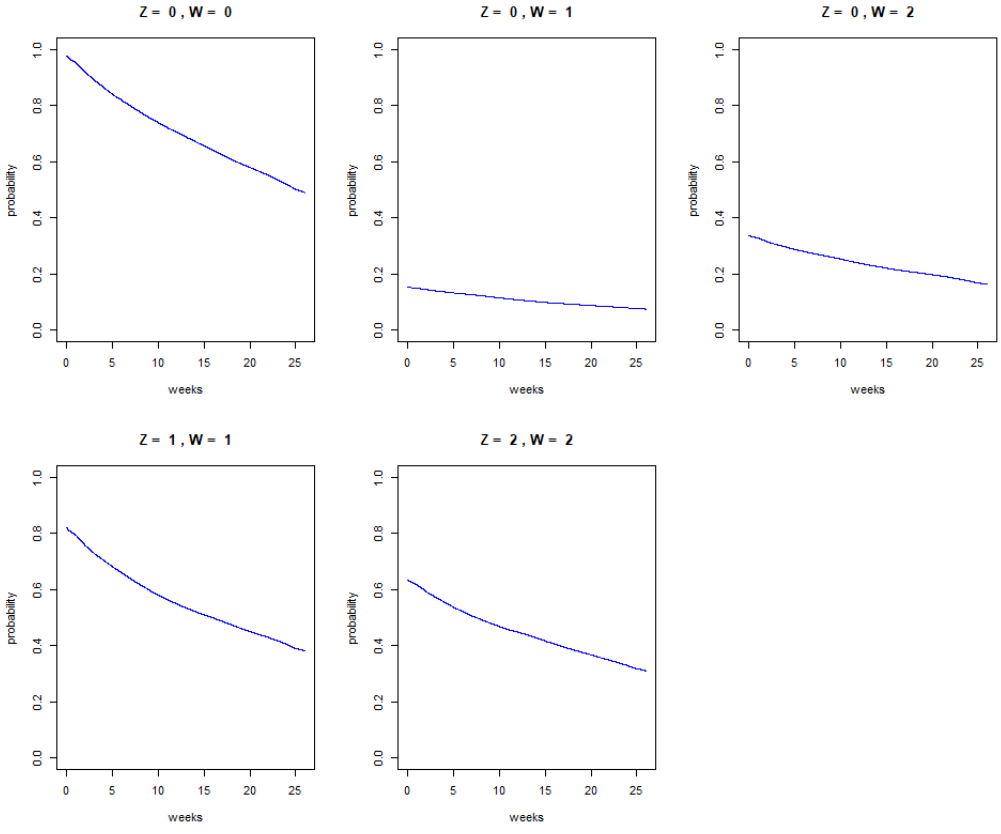

Figure 4.3: Survival function estimation for the full year duration

We minimize $\|A(\theta, \hat{S})\|_2^2$ (see Equation (3.15)), because the treatment variable has as many levels as the instrumental variable and the estimator is thus independent from V. Furthermore, φ is estimated for u ranging from 0.01 to 1, using a step size of 0.01 and the number of times that starting points Q are generated is set to 20. Figure (4.4) displays the value of the objective function for each duration and it starts to increase after u reaches 0.83 and 0.71 respectively for the first spell and the full year duration. The value of the objective function is also far from zero when u is close to zero, which can be explained by the fact that kernel smoothing is not performing well near boundaries. As shown by Beyhum et al. (2021), large values for the objective function indicate that the estimator is not identified. Therefore, the QTEs are only computed for u lower or equal to 0.83 and 0.71 respectively.

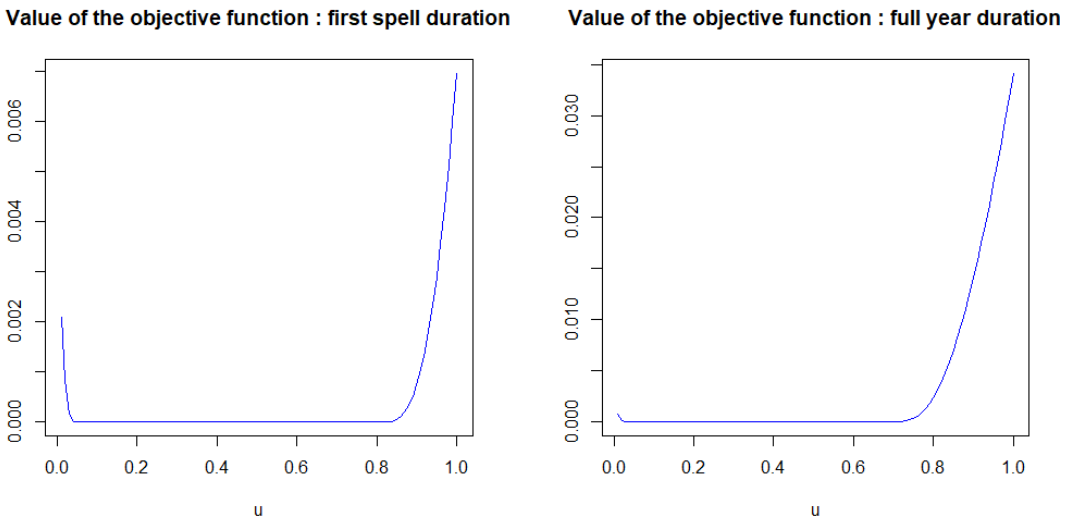

Figure 4.4: Value of the objective function for each duration

For each value of u we build a 95% confidence interval for the treatment effects, using the percentile bootstrap method with 1000 bootstrapped data sets. When the 95% confidence interval does not contain zero, the treatment effect is significant. The results for the first spell and the full year duration are respectively reported in Figures (4.5) and (4.6) and these are consistent with the findings of previous studies. Indeed, both treatments reduce the unemployment duration, whatever duration is considered. Furthermore, the JSIE treatment has a stronger effect than the HIE treatment and is significant for both durations. On the other hand, the HIE treatment is not significant for most values of u if the full year duration is considered, but it is significant if the first spell duration is considered. This suggests that people tend to return to unemployment within a year after they find a job due to HIE treatment. It would be interesting for future research to understand why this is so. Is the employer only interested in the bonus when the employee is hired, so that the employer will prefer an individual less suited for the job, but that will give him a cash bonus (and the employee is fired when the bonus is obtained)? Another possible explanation is that the employee tends to resign for whatever reason related to HIE experiment. We suggest to compare the ratio of people who resign or are fired in this subgroup compared to the whole population in order to gain insight into this problem. Unfortunately, this kind of data is not available in this experiment.

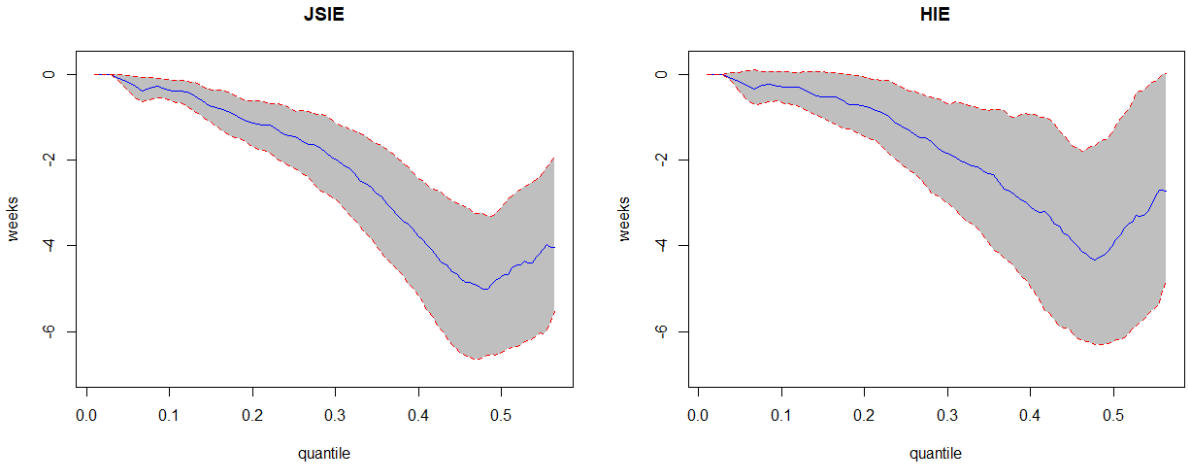

Figure 4.5: Confidence intervals of QTEs for the first spell duration

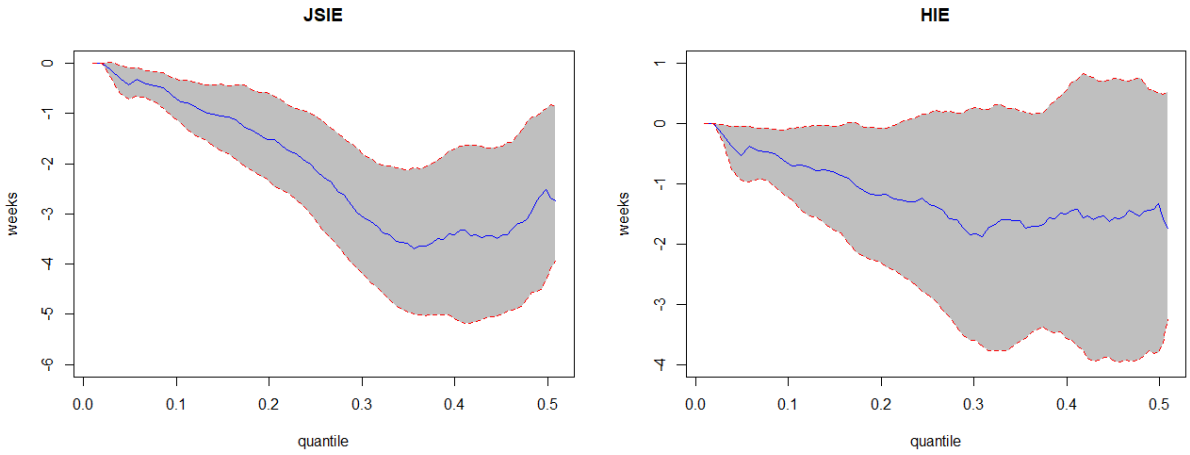

Figure 4.6: Confidence intervals of QTEs for the full year duration

Next, the quantile treatment effects for different subgroups of the data set will be analyzed in order to check whether or not the treatments have different effects on these different subgroups. The data set is split in different subgroups by means of covariates. For this study, the analysis is restricted to the full year duration in order to keep a clear overview and the number of bootstrapped data sets is limited to 100 (otherwise it takes too long to run the code).

4.2.1 Ethnicity

In this section, QTEs on different ethnic groups will be studied. In the data set 3136 participants ($\approx 25.91\%$) are Black, 913 participants ($\approx 7.54\%$) are Hispanic and 7785 participants ($\approx 64.33\%$) are White. Thus, approximately 98% of the participants are either Black, Hispanic or White, which makes that only a very small part of the data set consists of participants of other ethnic groups. For this reason, only the ethnic groups Black, Hispanic and White will be considered in this analysis. The covariate "Black" is defined as:

$$Black = \begin{cases} 1 \text{ if the participant is Black} \\ 0 \text{ if the participant is not Black.} \end{cases}$$

The computing time of the analysis in the statistical software environment R increases greatly when covariates are added to the model. To avoid this issue, the following equivalent estimation method is used. First the data set is split into two subsets. The first subset consists of all the observations for whom "Black" = 0, while the other subset consists of all the observations for whom "Black" = 1. Afterwards, QTEs will be calculated separately for each subsample. Note that discrete and continuous covariates need to be transformed to categorical variables before they can be used to split the data set. In addition, Equation (3.24) is used instead of Equation (3.11) when the data set is split.

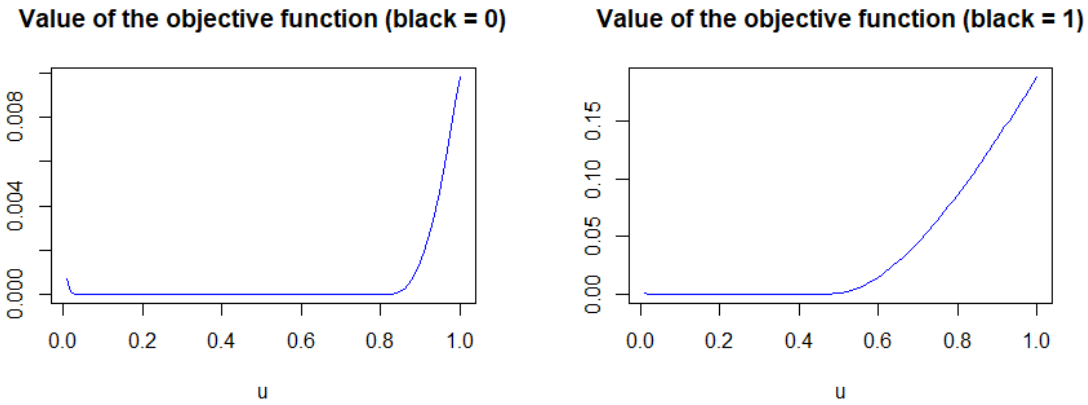

Figure 4.7: Value of the objective functions

In Figure (4.7), it is shown that the objective function generates low values for $0 < u < 0.90$ when "Black" equals zero and for $0 < u < 0.5$ when "Black" equals one. As mentioned in Chapter 3, the quantiles of the quantile regression function $\varphi(u)$ are represented by $1 - e^{-u}$ and the QTEs are derived from this function. Thus, the QTEs for the first subgroup ("Black" = 0) can be estimated reliably for every quantile between

0 ($= 1 - e^{-0}$) and 0.59 ($\approx 1 - e^{-0.90}$), while for the second subgroup ("Black" = 1), the QTEs can be estimated reliably only for every quantile between 0 ($1 - e^{-0}$) and 0.39 ($\approx 1 - e^{-0.5}$). For the first subgroup the QTEs can only be computed for quantiles smaller than 0.59, because approximately 41% of the observations of this subset are censored. Analogously, QTEs for the second subgroup can only be computed reliably for quantiles smaller than 0.39, because approximately 61% of the observations of this subset are censored.

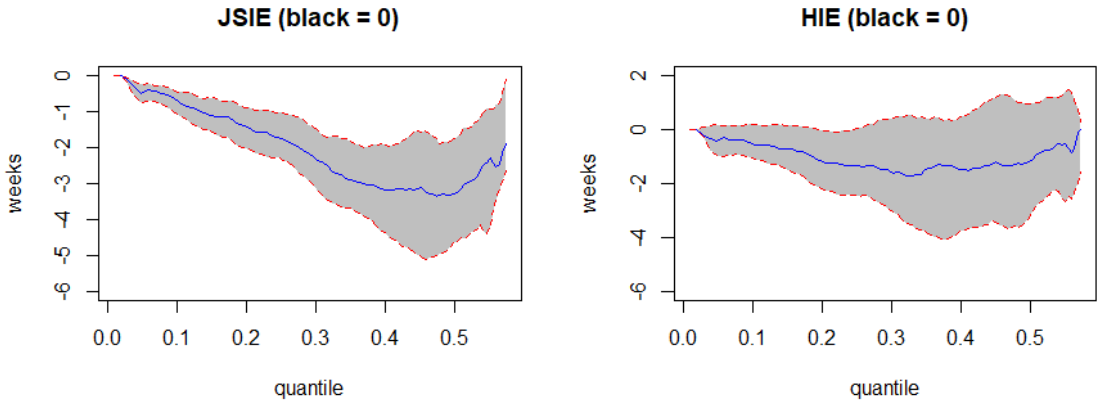

Figure 4.8: Confidence intervals of QTEs for Black = 0

A QTE is significant when the 95% confidence interval of the QTE does not contain zero. Figure (4.8) shows that the JSIE treatment has a significant effect on the unemployment duration of participants for whom Black is equal to zero, while the HIE treatment does not have a significant impact on these participants.

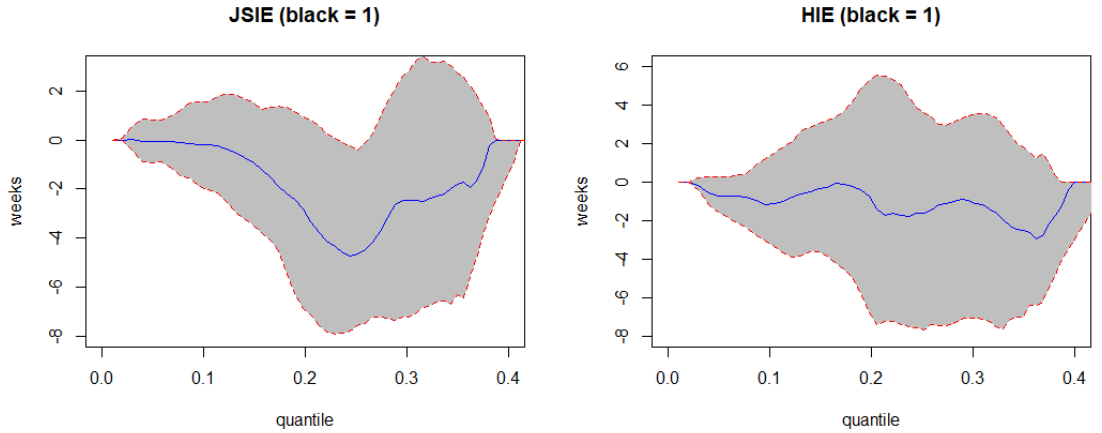

Figure 4.9: Confidence intervals of QTEs for Black = 1

It is clear from Figure (4.9) that both treatments do not have a significant effect for any quantile of participants for whom "Black" is equal to one. While this is not surprising for the HIE treatment, it is noteworthy that also the JSIE treatment does not have a significant effect on any quantile (except for some quantiles around 0.25) for these participants. This finding suggests that Black participants get less incentive to find a job from receiving a cash bonus than the other participants.

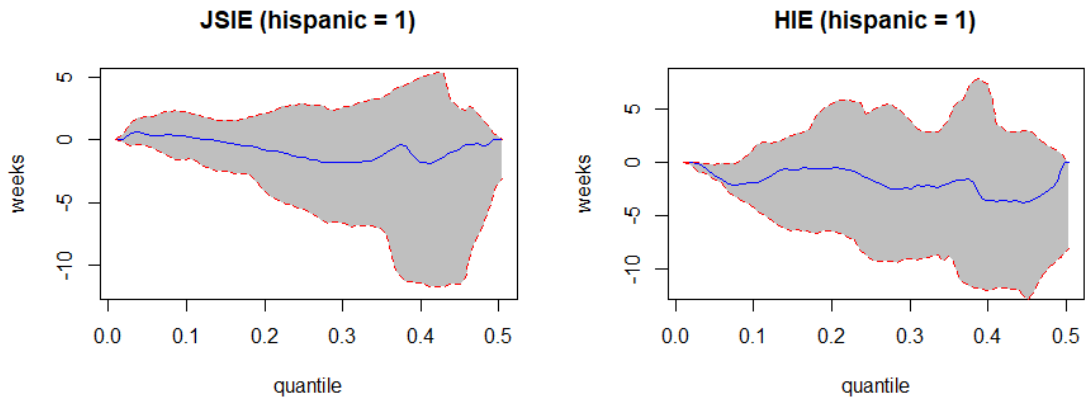

Figure 4.10: Confidence intervals of QTEs for Hispanic = 1

In Figure (4.10), it is shown that the JSIE treatment also does not have a significant effect on participants that are Hispanic. This indicates that the JSIE treatment only has a significant effect on participants for whom "White" is equal to one, which is confirmed by Figure (4.11) below. This figure also shows that the HIE treatment is significant for some quantiles of this subgroup. All of this suggests that ethnicity has a significant impact on the effectiveness of such bonuses.

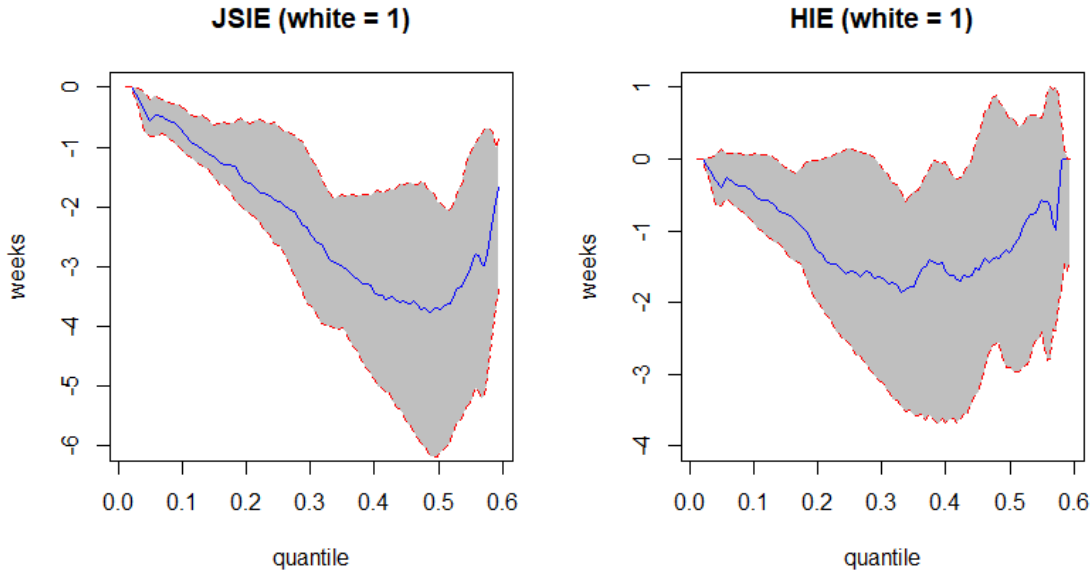

Figure 4.11: Confidence intervals of QTEs for White = 1

4.2.2 Gender

The covariate of interest in this section is "male", which is equal to zero when the participant is female and is equal to one when the participant is male. In the data set, 6650 participants are male and 5451 participants are female.

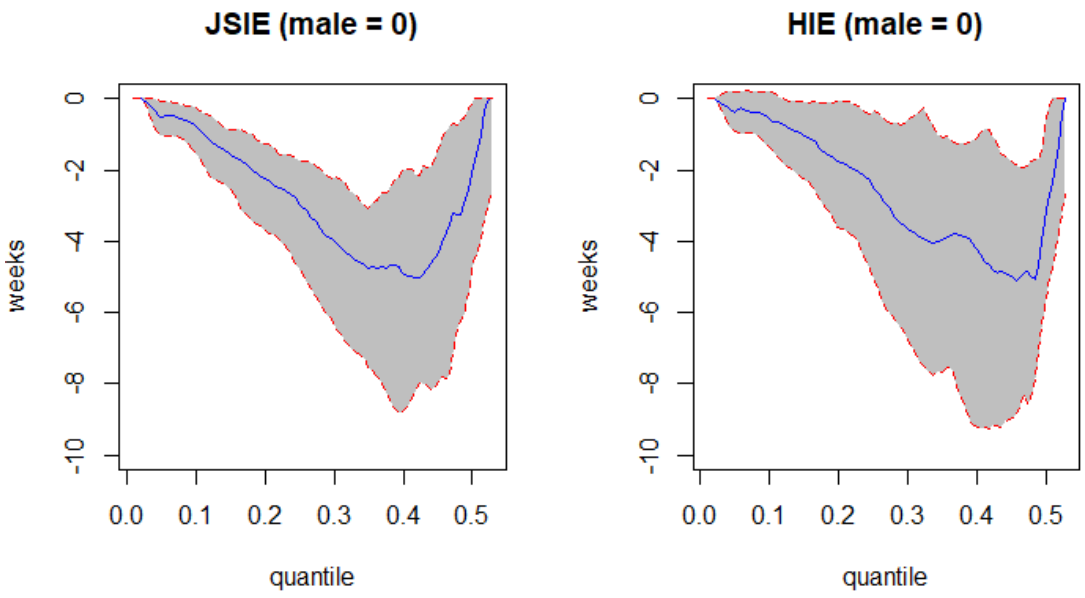

Figure 4.12: Confidence intervals of QTEs for women

Figure (4.12) shows that both treatments have a significant effect for most quantiles for females. The JSIE treatment reduces the unemployment duration of females significantly for all quantiles, while the HIE treatment does reduce this duration significantly for all quantiles larger than 0.1. Furthermore, Figure (4.13) shows that for the male participants the HIE treatment is not significant for any quantile, while the JSIE treatment is only significant for quantiles larger than 0.2. It is interesting to see that the HIE treatment has a much bigger effect on females than on males, especially on females for whom it takes more time than average to find a job. All of this suggests that gender has a significant impact on the effectiveness of such bonuses.

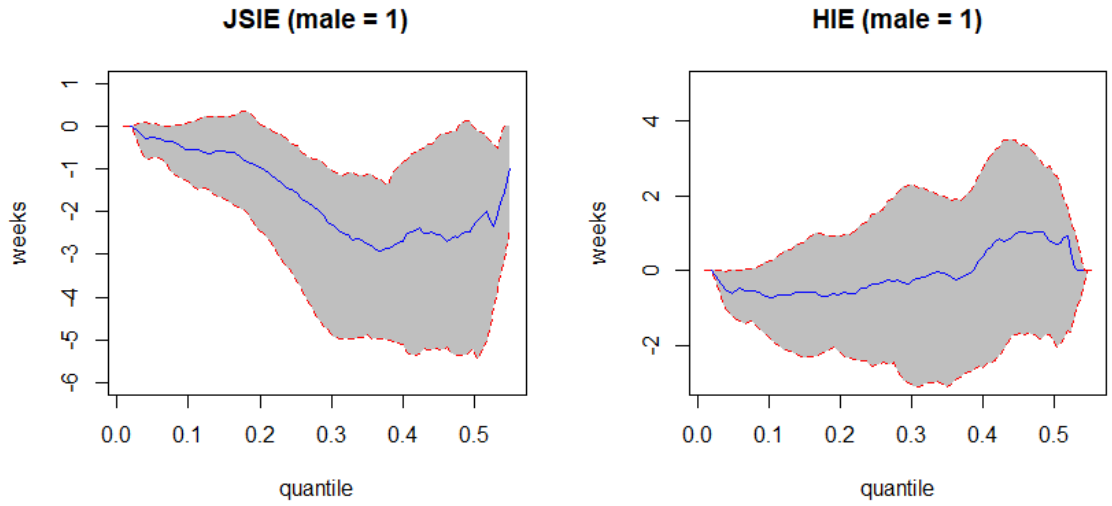

Figure 4.13: Confidence intervals of QTEs for men

4.2.3 Age

The covariate that will be analyzed in this section is the variable "age", which represents the age (in years) of the participants. We transform this discrete variable into a categorical variable by splitting it into four parts (using quartiles), which ensures that every part has an equal amount of observations. Having enough observations for each variable is crucial to have accurate estimations and it is even more important in the context of censoring. Moreover, splitting in four different categories is relevant to have interpretable results. With more categories, one would find different results for more subgroups, but so many categories would not lead to a comprehensible picture of the outcomes. With less categories, one could miss possible heterogeneity and would not get a comprehensive picture of the outcomes. These four quarters represent the categories of the new categorical variable (AGE_{cat}) as follows:

$$AGE_{cat} = \begin{cases} 3 \text{ if } 39 \leq \text{age} \\ 2 \text{ if } 31 \leq \text{age} < 39 \\ 1 \text{ if } 26 \leq \text{age} < 31 \\ 0 \text{ if age} < 26 \end{cases}$$

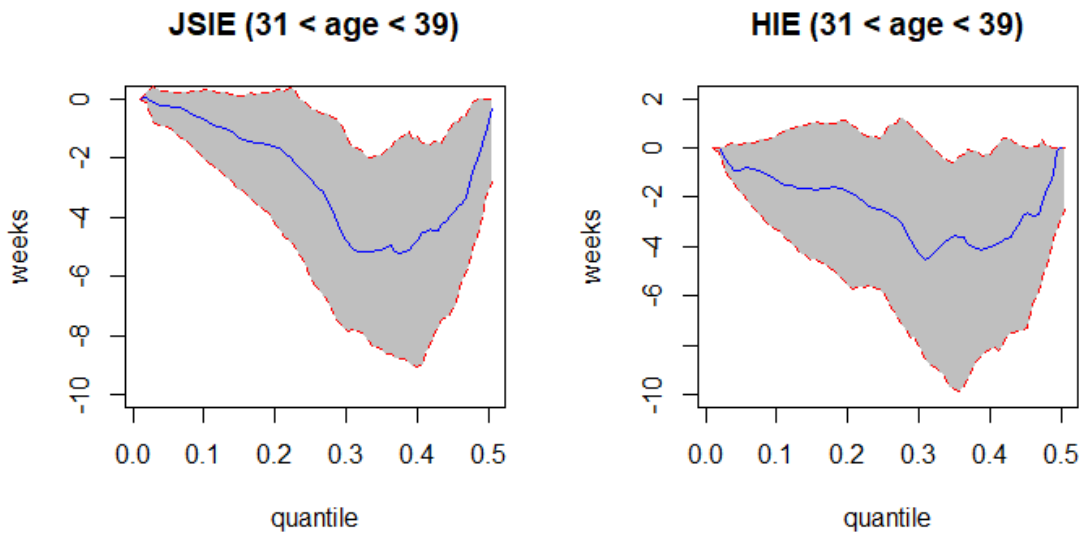

Figure 4.14: Confidence intervals of QTEs for $AGE_{cat} = 2$

The QTEs are represented by Figures (4.14), (4.15), (A.2) and (A.3) (the last two figures can be found in Appendix A). It can be seen that the JSIE treatment is significant for some quantiles of all the subgroups. However, this treatment effect seems to be the strongest for the subgroup with participants with $AGE_{cat} = 2$ (see Figure (4.14)). Another interesting observation is that for this same subgroup, the HIE treatment is significant for the quantiles between 0.3 and 0.4, while this treatment is not significant for any quantile of the other subgroups. This indicates that people in their thirties benefit the most from both treatments. Once again, we see that adding a covariate about age can modify the conclusion of that kind of experiments.

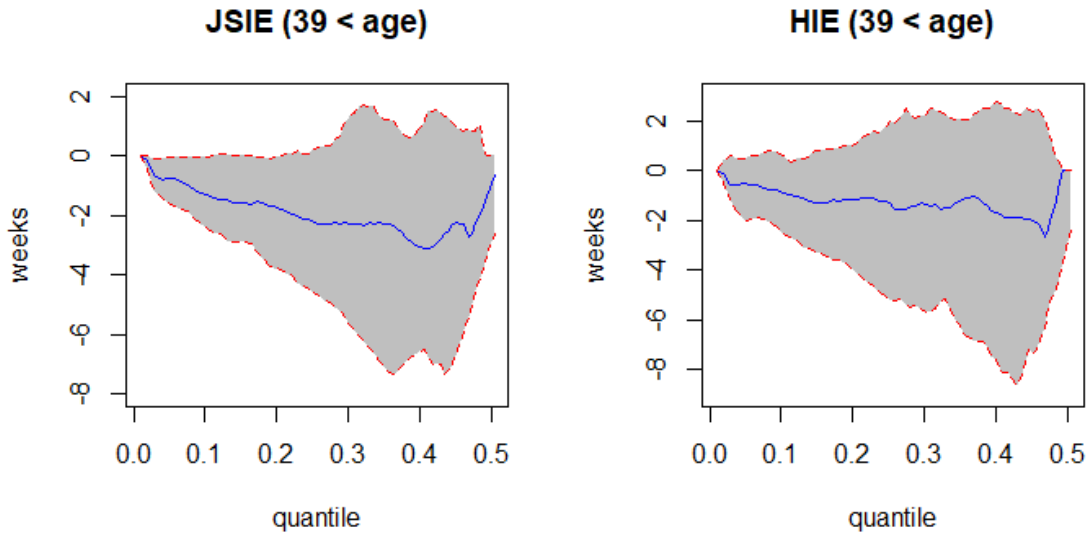

Figure 4.15: Confidence intervals of QTEs for $AGE_{cat} = 3$

4.2.4 Earnings before, during and after unemployment

The first covariate that will be analyzed in this section is the variable "avprearn", which represents the average quarterly earnings in the first four of the five completed quarters prior to the filing for unemployment insurance (UI). Just like in the previous section, the variable "avprearn" is split into four quarters using quartiles. These four quarters represent the categories of the new categorical variable ($AVPREARN_{cat}$) as follows:

$$AVPREARN_{cat} = \begin{cases} 3 \text{ if } 4235 \leq \text{avprearn} \\ 2 \text{ if } 2700 \leq \text{avprearn} < 4235 \\ 1 \text{ if } 1529 \leq \text{avprearn} < 2700 \\ 0 \text{ if avprearn} < 1529 \end{cases}$$

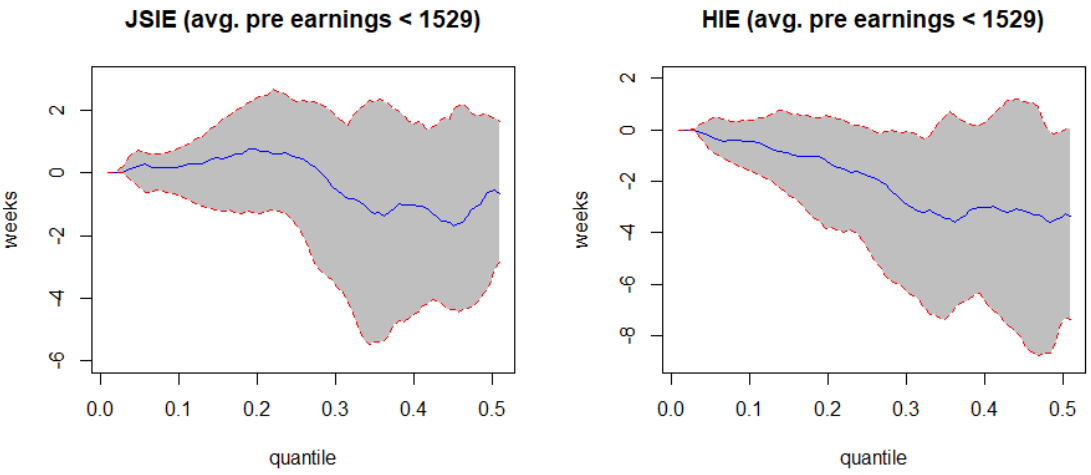

Figure 4.16: Confidence intervals of QTEs for $AVPREARN_{cat} = 0$

In Figure (4.16), it is shown that the JSIE treatment does not have a significant effect on the unemployment duration of participants with $AVPREARN_{cat} = 0$, while the HIE treatment effect is much stronger for this subgroup. In contrast, it is clear from Figure (4.17) that the JSIE treatment does reduce the duration of unemployment of participants with $AVPREARN_{cat} = 1$ significantly for all quantiles smaller than 0.45. Furthermore, Figure (A.4) (see Appendix A) indicates that the JSIE treatment has a significant effect on the participants with $AVPREARN_{cat} = 2$ for the quantiles smaller than 0.4 and even the HIE treatment is slightly significant for some quantiles of this subgroup. Finally, Figure (A.5) (see Appendix A) shows that the JSIE treatment is significant for the subgroup with $AVPREARN_{cat} = 3$ for most quantiles bigger than 0.1, while the HIE treatment is not significant at all for this subgroup. It is noteworthy that the JSIE treatment effect is not significant for participants with a low income before filing for unemployment (see Figure (4.16)), because the $500 bonus could make a big difference for participants of this subgroup. As mentioned in the introduction (see Chapter 1), Card and Krueger (2000) showed that an increase in the minimum wage does not necessarily imply an increase in unemployment. It could be interesting to examine to what extend this can be linked with our findings. It is likely that financial incentives are not the key reasons why very poor people stay unemployed.

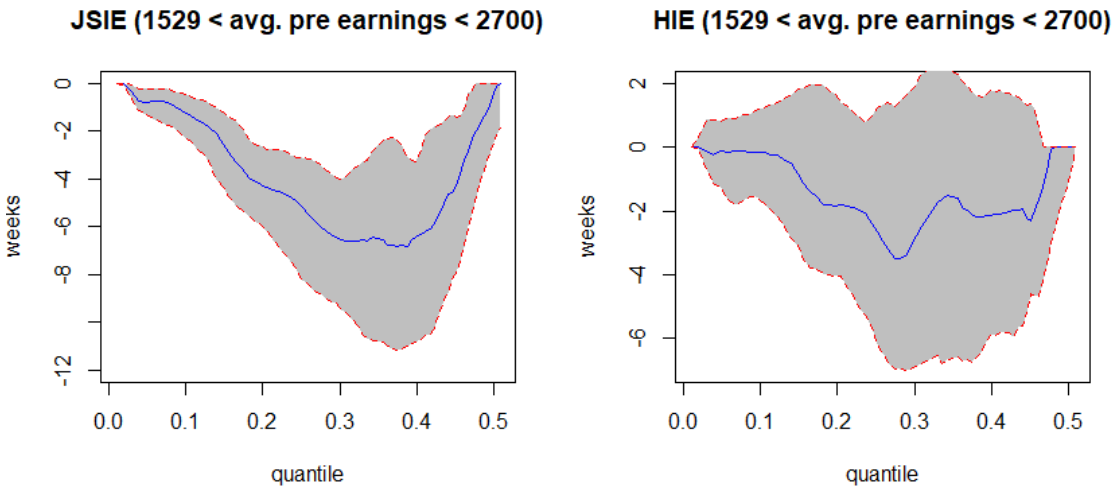

Figure 4.17: Confidence intervals of QTEs for $AVPREARN_{cat} = 1$

Next, the covariate "pospearn" will be discussed. This variable represents the quarterly earnings of the claimant in the first post-claim quarter in which the claimant had earnings. First all the observations for whom "pospearn" $= 0$ are removed from the data set. Afterwards the continuous variable is split by the quartiles in the categorical variable $POSPEARN_{cat}$.

$$POSPEARN_{cat} = \begin{cases} 3 \text{ if } 3135 \leq \text{pospearn} \\ 2 \text{ if } 1578 \leq \text{pospearn} < 3135 \\ 1 \text{ if } 595 \leq \text{pospearn} < 1578 \\ 0 \text{ if pospearn} < 595 \end{cases}$$

It is not suprprising that the subsets with lower values for "pospearn" have much more censored values than the subsets with high values for this variable. For this reason the QTEs of the subsets with higher values for "pospearn" can be calculated for more quantiles than the QTEs of the subsets with low values for this covariate.

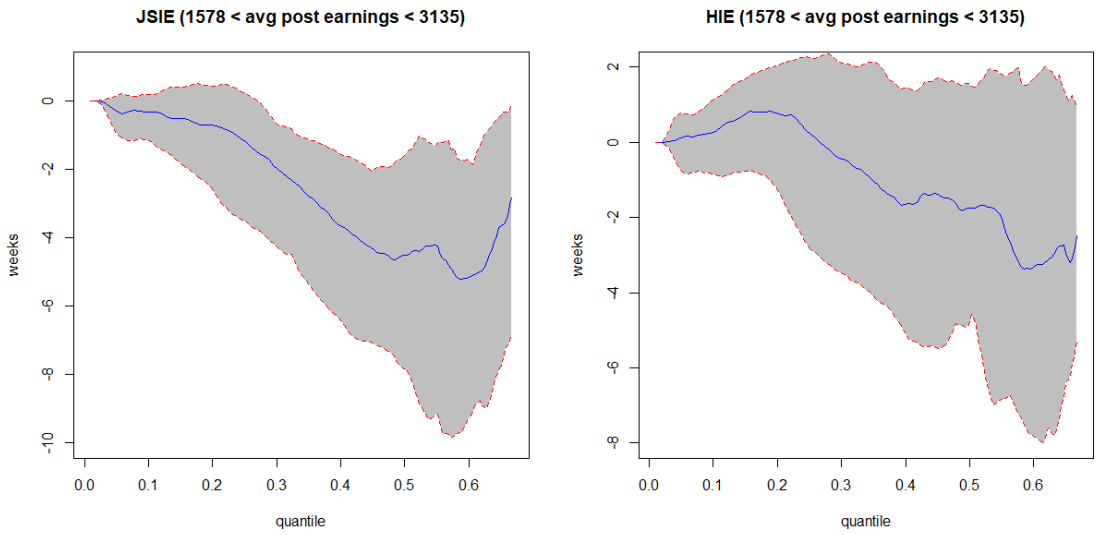

Figure 4.18: Confidence intervals of QTEs for $POSPEARN_{cat} = 2$

The JSIE treatment does reduce the unemployment duration for all subgroups, except for the subgroup with the lowest earnings in the first post-claim quarter (see Figure (A.6)). For the subgroup with $POSPEARN_{cat} = 2$ the JSIE treatment is significant for all quantiles larger than 0.3 (see Figure (4.18)), while for the subgroup with $POSPEARN_{cat} = 3$ this treatment is significant for all quantiles smaller than 0.6 (see Figure (4.19)). Furthermore, it is shown in Figure (A.7) (see Appendix A) that the JSIE treatment reduces the unemployment duration of participants with $POSPEARN_{cat} = 1$ significantly for some quantiles between 0.3 and 0.4. All this suggests that the JSIE treatment is more effective for participants with high earnings in the first post-claim quarter.

The HIE treatment is only significant for some quantiles smaller than 0.3 of the subgroup with $POSPEARN_{cat} = 3$. It is interesting to see that the HIE treatment reduces the unemployment duration for participants of this subgroup for whom it takes less time than average to find a job, while the unemployment duration of participants of this subgroup for whom it takes more time than average to find a job is (not significantly) increased by this treatment.

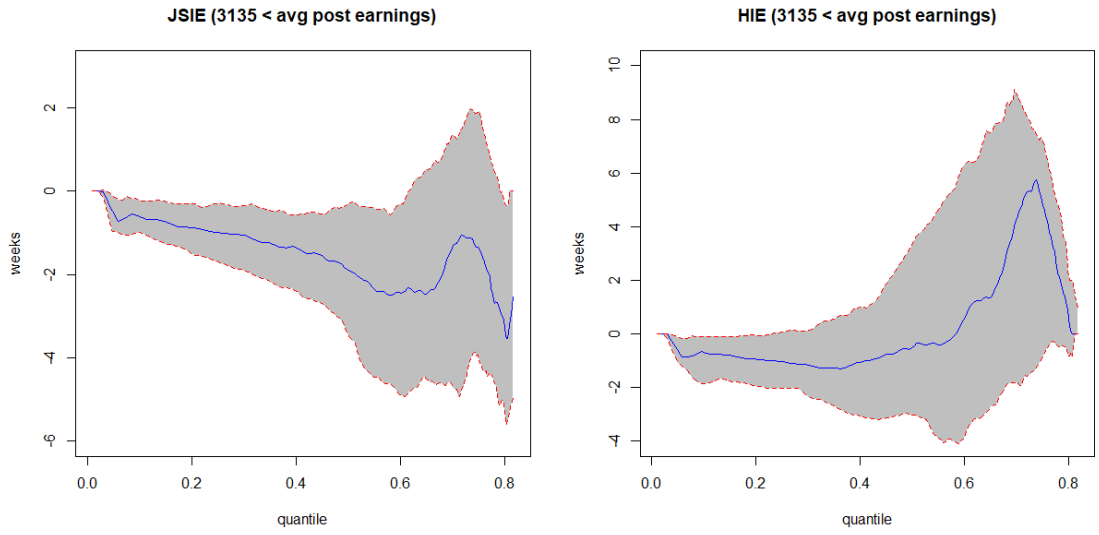

Figure 4.19: Confidence intervals of QTEs for $POSPEARN_{cat} = 3$

Finally the covariate "wkbeneft", which represents the weekly benefit amount during the enrollment period, will be added to the original model. After transforming this continuous variable, the categorical variable $WKBENEFT_{cat}$ is obtained. The highest amount of weekly benefits is €161 and there are 4122 observations for whom the weekly benefit during unemployment is equal to this value. Consequently the 4^{th} quarter of this variable only consists of values equal to 161.

$$WKBENEFT_{cat} = \begin{cases} 3 \text{ if wkbeneft} = 161 \\ 2 \text{ if } 126 \leq \text{wkbeneft} < 161 \\ 1 \text{ if } 83 \leq \text{wkbeneft} < 126 \\ 0 \text{ if wkbeneft} < 83 \end{cases}$$

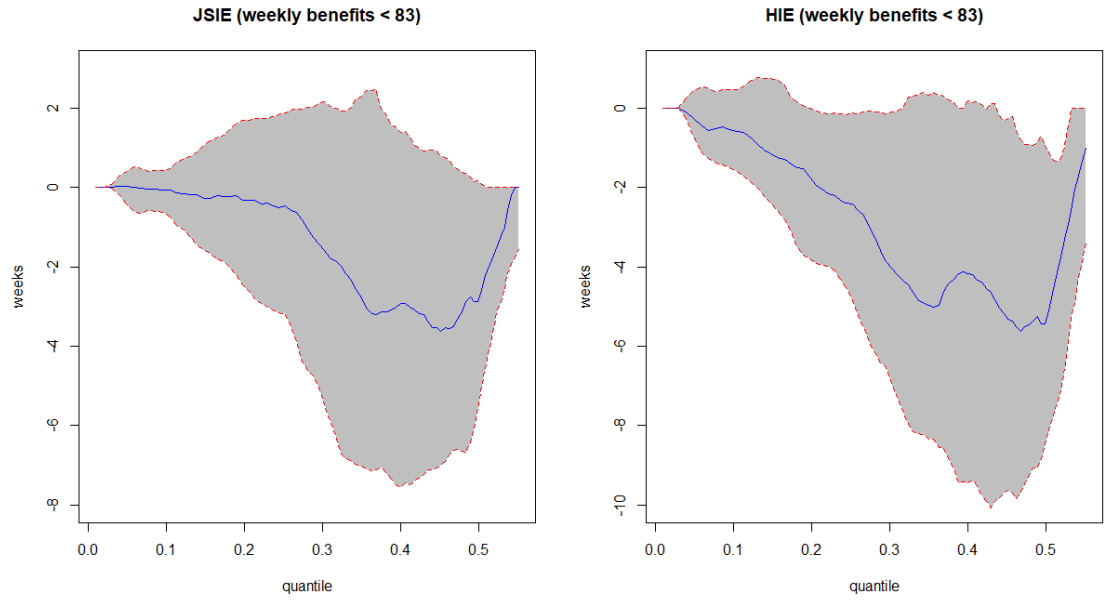

Figure 4.20: Confidence intervals of QTEs for $WKBENEFT_{cat} = 0$

Figure (4.20) shows that the JSIE treatment is not significant for any quantile of participants with the lowest weakly benefits, while the unemployment duration of participants with the highest weekly benefits is reduced significantly for all quantiles by this treatment (see Figure (4.21)). In Figure (A.8) it can be seen that for participants with $WKBENEFT_{cat} = 1$ this treatment reduces the duration significantly for all quantiles between 0.2 and 0.4, while this treatment is only significant for some quantiles around 0.25 for participants with $WKBENEFT_{cat} = 2$ (see Figure A.9). These findings suggest that the JSIE treatment is most effective for participants with high weekly benefits, while it is not effective for participants with low weekly benefits.

It is also interesting to see that the HIE treatment is only significantly reducing the unemployment duration of the participants with the lowest weekly benefits, especially for the higher quantiles of this subgroup. This suggests that participants who get a low amount of weekly benefits and who take a long time to find a job, benefit the most from the HIE treatment.

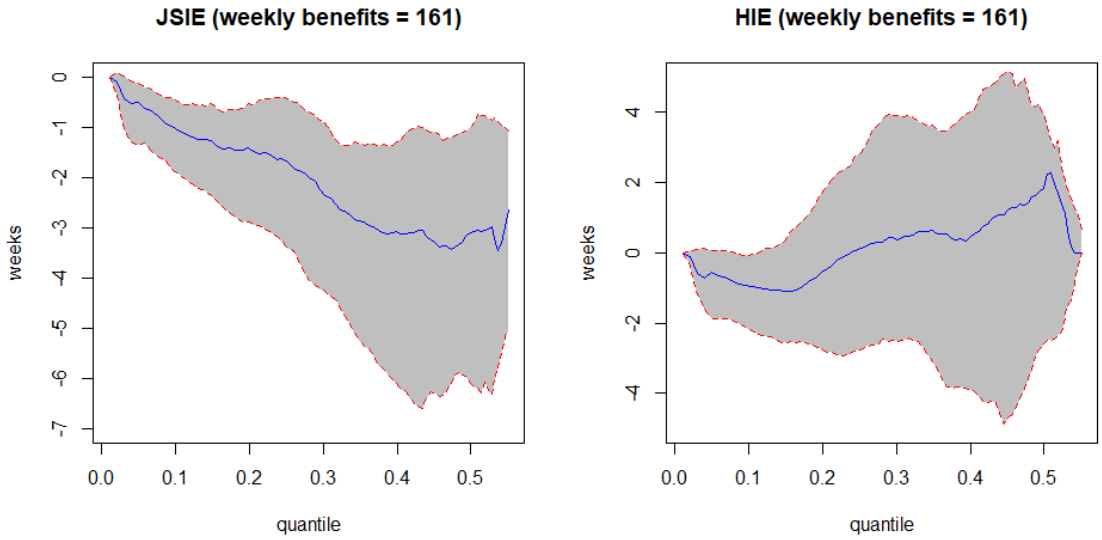

Figure 4.21: Confidence intervals of QTEs for $WKBENEFT_{cat} = 3$

Chapter 5

Conclusion

5.1 Summary of the results

Both treatments reduce the unemployment duration, whatever type of duration is considered. The JSIE treatment has a stronger effect than the HIE treatment and is significant for both durations. The HIE treatment is not significant if the full year duration is considered, while it is significant if the first spell duration is considered. This suggests that people tend to return to unemployment within a year after they find a job due to HIE treatment. Furthermore, the treatments have different effects on different subgroups of the data set. Firstly, ethnicity has a significant impact on the effectiveness of such bonuses. The JSIE treatment is indeed only significant for White people. Moreover, gender also influences the efficiency of the treatments. The HIE treatment is significant for females, while it is not for males. Furthermore, age also has an impact on the results. Our findings suggest that people in their thirties get more incentive to find a job from receiving a cash bonus and the HIE treatment effect is also most pronounced in this subgroup. Lastly, earnings of an individual have a significant impact on the conclusions we can draw. The JSIE treatment effect is not significant for participants with a low income before filing for unemployment, which is surprising, because the bonus could make a big difference for participants of this subgroup. The HIE treatment is only significantly reducing the unemployment time of the participants with the lowest weekly benefits, especially for the higher quantiles of this subgroup. This suggests that participants who get a low amount of weekly benefits and who take a long time to find a job, benefit the most from the HIE treatment.

5.2 Limits and future research

It was stated several times that people tend to return to unemployment within a year after they find a job due to HIE treatment. It would be interesting for future research to understand why this is so. Is the employer only interested in the bonus when the individual is hired, so that the employer will prefer an individual less suited for the job but that will give him a bonus (and the individual is fired when he benefited from the bonus)? Does the employee tend to resign for whatever reasons related to HIE experiment? We suggest to compare the ratio of people who resign/are fired in this subgroup compared to the whole population to gain insight about this problem. Unfortunately, this kind of data is not available in this experiment.

Another interesting study would be to investigate the influence between explanatory variables. For instance, findings suggest that cash bonuses are less efficient for Black and Hispanic people but also for people with a low income. One could argue that those

two results are correlated because the explanatory variables are correlated. It would be therefore interesting to investigate the impact of explanatory variables conditionally to the others.

Moreover, as mentioned in Section 4.2.4, It could be interesting to examine to what extend the fact that the JSIE treatment is not significant for participants with a low income could be linked to the findings of Card and Krueger (2000), showing that an increase in the minimum wage does not necessarily imply an increase in unemployment. It is likely that financial incentives are not the key reasons why very poor people stay unemployed.

Finally, as mentioned in the introduction, the framework proposed by Wei et al. (2021) is also suited for our empirical application. It could therefore be interesting to apply this method to the data set and to enhance the model in order to include covariates. A comprehensive comparison of the results between the two models could then be conducted.

List of Figures

2.1 Ordered sample divided in quantiles . . . 8
2.2 Asymmetric loss function $\rho_\tau(u)$ (Koenker, 2005, Chapter 1) . . . 9
2.3 Causal relationship between W, X, U and Y . . . 13

4.1 Kaplan-Meier estimation for the first spell duration . . . 24
4.2 Survival function estimation for the first spell duration . . . 25
4.3 Survival function estimation for the full year duration . . . 25
4.4 Value of the objective function for each duration . . . 26
4.5 Confidence intervals of QTEs for the first spell duration . . . 27
4.6 Confidence intervals of QTEs for the full year duration . . . 27
4.7 Value of the objective functions . . . 28
4.8 Confidence intervals of QTEs for Black = 0 . . . 29
4.9 Confidence intervals of QTEs for Black = 1 . . . 29
4.10 Confidence intervals of QTEs for Hispanic = 1 . . . 30
4.11 Confidence intervals of QTEs for White = 1 . . . 30
4.12 Confidence intervals of QTEs for women . . . 31
4.13 Confidence intervals of QTEs for men . . . 31
4.14 Confidence intervals of QTEs for $AGE_{cat} = 2$ . . . 32
4.15 Confidence intervals of QTEs for $AGE_{cat} = 3$ . . . 33
4.16 Confidence intervals of QTEs for $AVPREARN_{cat} = 0$ . . . 33
4.17 Confidence intervals of QTEs for $AVPREARN_{cat} = 1$ . . . 34
4.18 Confidence intervals of QTEs for $POSPEARN_{cat} = 2$ . . . 35
4.19 Confidence intervals of QTEs for $POSPEARN_{cat} = 3$ . . . 36
4.20 Confidence intervals of QTEs for $WKBENEFT_{cat} = 0$ . . . 36
4.21 Confidence intervals of QTEs for $WKBENEFT_{cat} = 3$ . . . 37

A.1 Kaplan-Meier estimation for full year duration . . . 42
A.2 Confidence intervals of QTEs for $AGE_{cat} = 0$ . . . 43
A.3 Confidence intervals of QTEs for $AGE_{cat} = 1$ . . . 43
A.4 Confidence intervals of QTEs for $X_{categorical} = 2$ . . . 44
A.5 Confidence intervals of QTEs for $X_{categorical} = 3$ . . . 44
A.6 Confidence intervals of QTEs for $POSPEARN_{cat} = 0$ . . . 45
A.7 Confidence intervals of QTEs for $POSPEARN_{cat} = 1$ . . . 45
A.8 Confidence intervals of QTEs for $WKBENEFT_{cat} = 1$ . . . 46
A.9 Confidence intervals of QTEs for $WKBENEFT_{cat} = 2$ . . . 46

List of Tables

2.1 Numerical example for kernel regression 11
2.2 Weights of the kernel regression 12
2.3 Numerical example for survival analysis 14

4.1 Mean of different variables for each group 21
4.2 Average unemployment duration (first spell) of each group 21
4.3 Average unemployment duration (full year) of each group 21
4.4 Sample sizes of the experiment 22

Appendix A

Figures

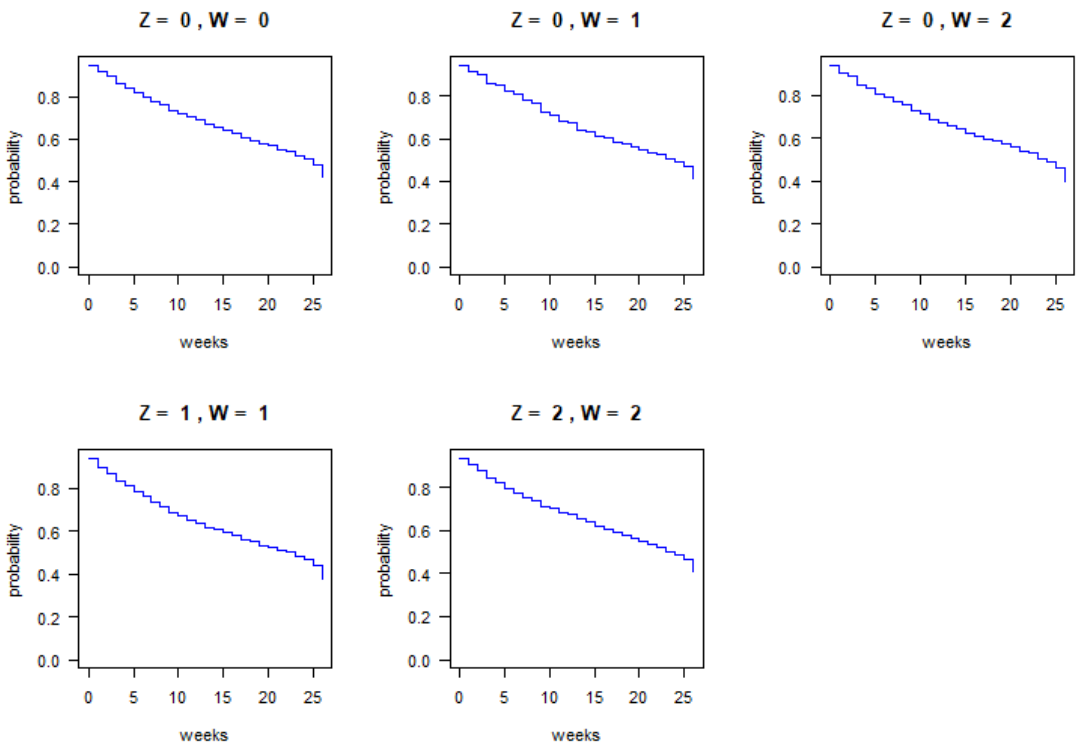

Figure A.1: Kaplan-Meier estimation for full year duration

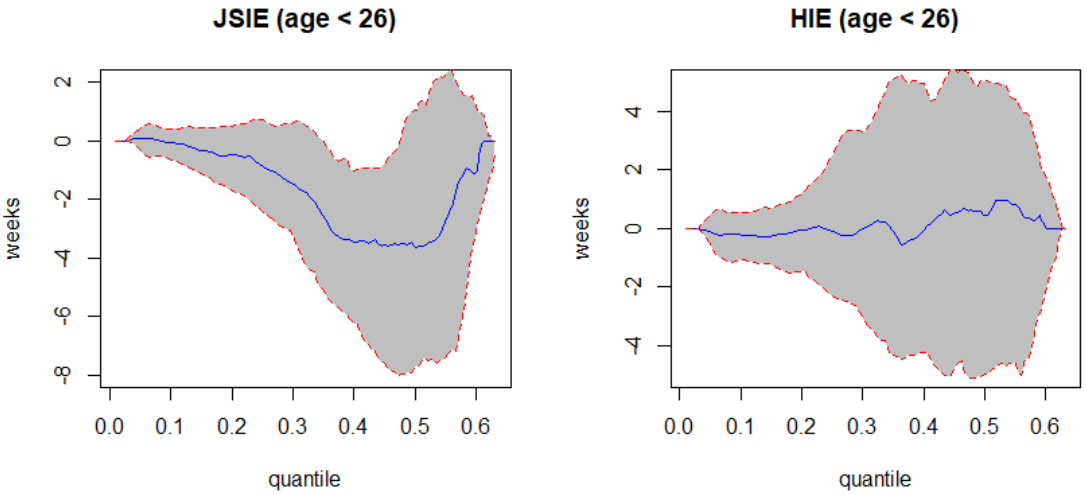

Figure A.2: Confidence intervals of QTEs for $AGE_{cat} = 0$

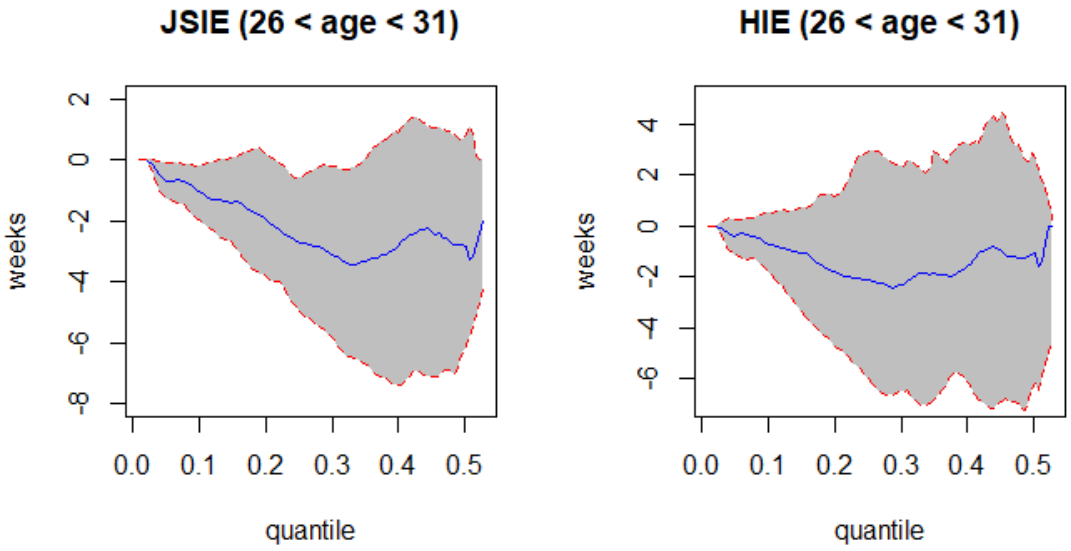

Figure A.3: Confidence intervals of QTEs for $AGE_{cat} = 1$

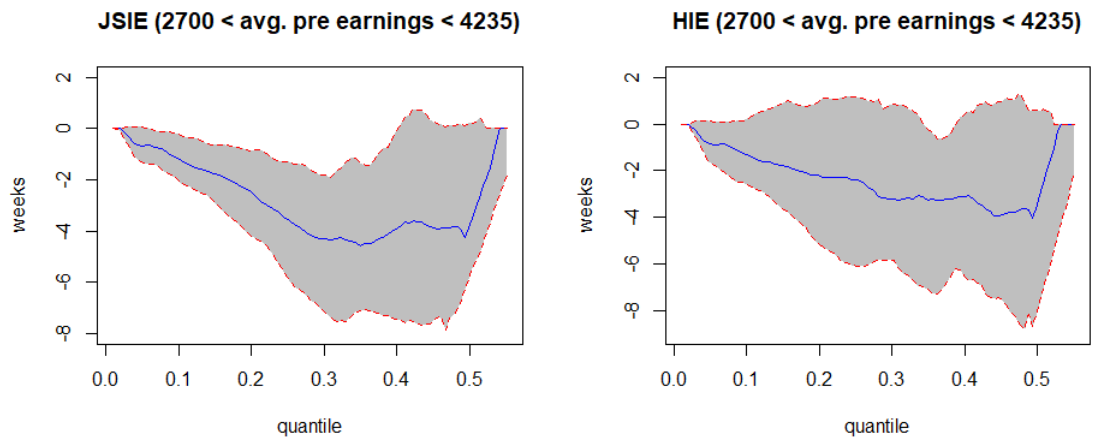

Figure A.4: Confidence intervals of QTEs for $X_{categorical} = 2$

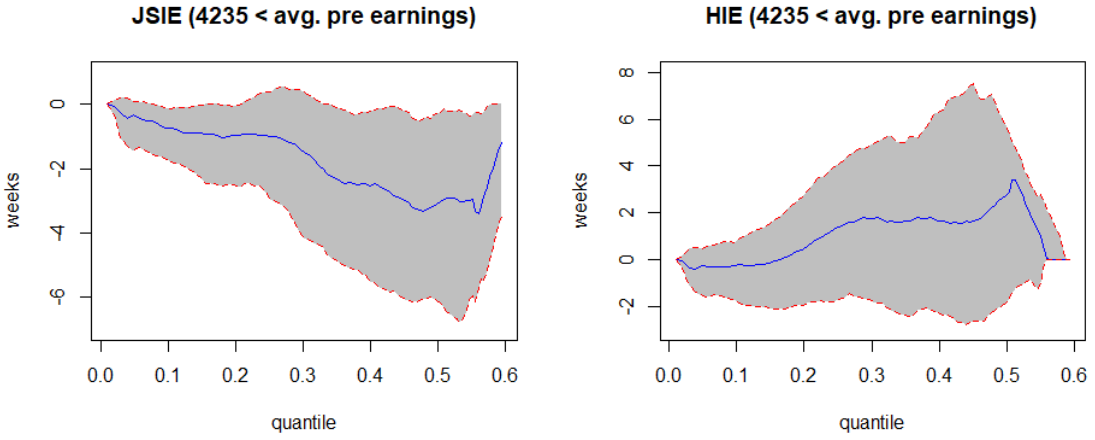

Figure A.5: Confidence intervals of QTEs for $X_{categorical} = 3$

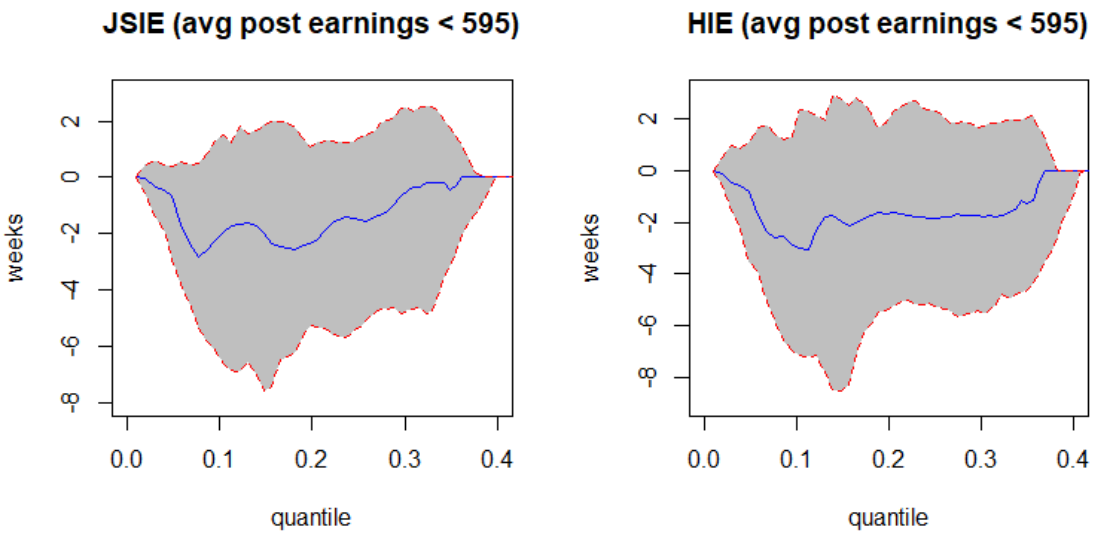

Figure A.6: Confidence intervals of QTEs for $POSPEARN_{cat} = 0$

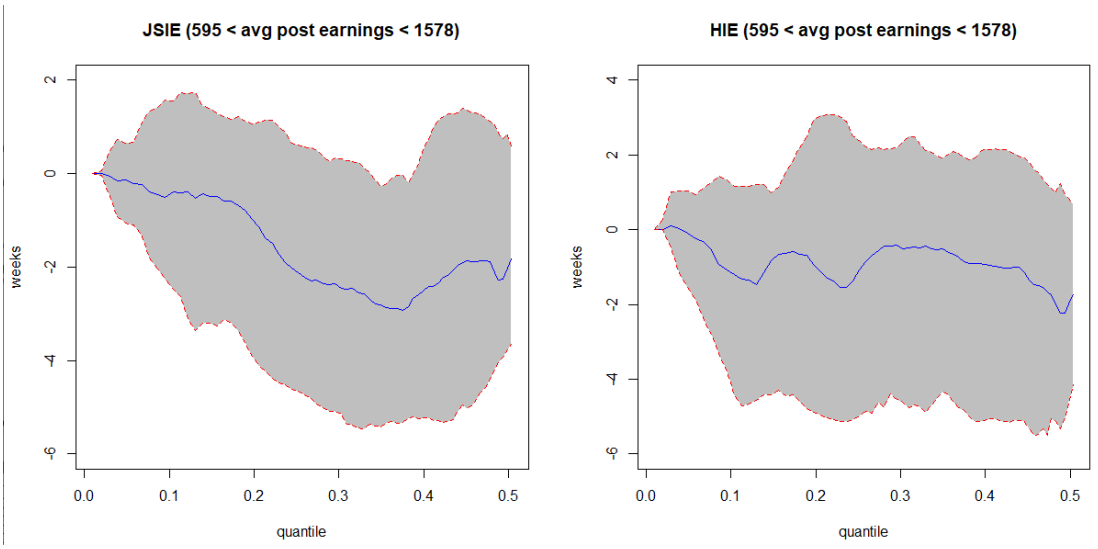

Figure A.7: Confidence intervals of QTEs for $POSPEARN_{cat} = 1$

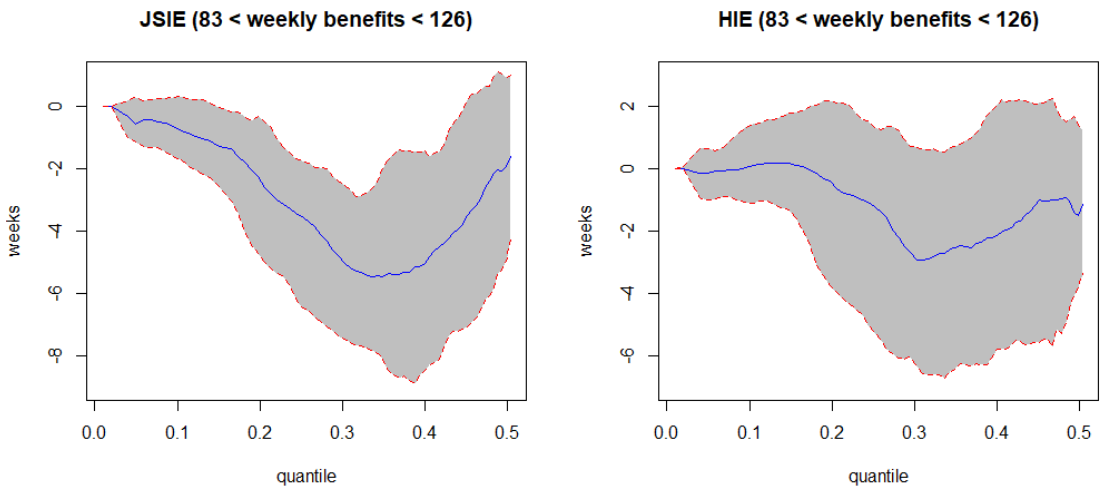

Figure A.8: Confidence intervals of QTEs for $WKBENEFT_{cat} = 1$

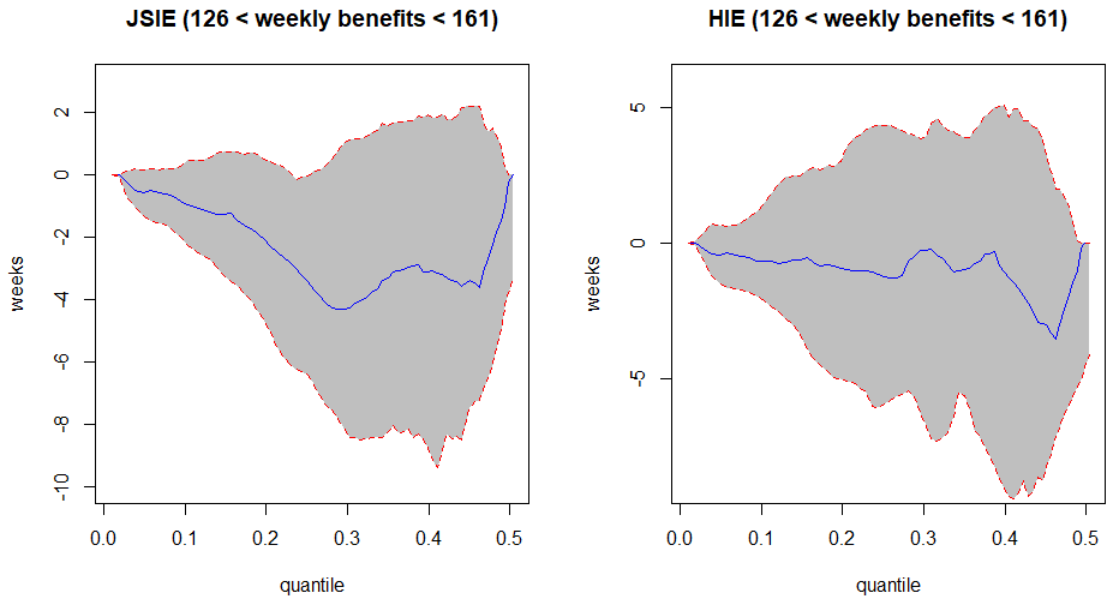

Figure A.9: Confidence intervals of QTEs for $WKBENEFT_{cat} = 2$

Appendix B

R code

B.1 The model and descriptive statistics

```
#Data preparation--------------------

#Loading libraries and importing data
{
  # Package names
  packages <-
    c("survival",
      "foreach",
      "quantreg",
      "haven",
      "doParallel",
      "KernSmooth",
      "xtable")

  # Install packages not yet installed
  installed_packages <- packages %in% rownames(installed.packages())
  if (any(installed_packages == FALSE)) {
    install.packages(packages[!installed_packages],
    repos='http://cran.us.r-project.org')
  }
  # Packages loading
  invisible(lapply(packages, library, character.only = TRUE))
} #LOAD PACKAGES


#import dataframe
illinois <- as.data.frame(read_dta("illinois_reemployment_bonus_experiment.dta"))

#Define duration : choose wkpaid for first spell or wkspdbye for full year of UI
wkpaid= as.numeric(illinois$wkspdbye)

#Transforming NA in 0 for the variable "agrees to participate"
illinois$lagree[is.na(illinois$lagree)] = 0

#Creating variables Z, W and Delta
Z = vector(mode = "numeric", length = nrow(illinois)) #Treatment variable
```

```
W = vector(mode = "numeric", length = nrow(illinois)) #Instrumental variable
delta = vector(mode = "numeric", length = nrow(illinois)) #Censoring indicator

for (i in 1:nrow(illinois)){
  if ((illinois$hie[i]==1)&(illinois$lagree[i]==1))
  {
    Z[i] = 2
  }
  if ((illinois$jsie[i]==1)&(illinois$lagree[i]==1))
  {
    Z[i] = 1
  }
  if ((illinois$hie[i]==1))
  {
    W[i] = 2
  }
  if ((illinois$jsie[i]==1))
  {
    W[i] = 1
  }
  if (illinois$wkpaid[i]<26)
  {
    delta[i] = 1
  }
}

   ############################
   # descriptive statistics
   ############################
 hist(wkpaid, breaks = max(wkpaid)+1)
 summary(wkpaid)
 table = matrix(nrow = 49, ncol = 3)
 rownames(table) = colnames(illinois)
 colnames(table) = c("Control group", "JSIE", "HIE")
 table[,1] = round(colMeans(illinois[W==0,]),3)
 table[,2] = round(colMeans(illinois[W==1,]),3)
 table[,3] = round(colMeans(illinois[W==2,]),3)
 table = table[c(1,22,6,46,15,24,27, 2),]
 rownames(table) = c("Age", "Male","Black", "White", "Hispanic",
 "Native American", "Others", "Base period earnings" )

 #this command gives the latex code for table 4
 print(xtable(table, type = "latex", digits = 3), file = "desc_stat.tex")

 #Average unemployment duration
 avg_dur = matrix(ncol = 8, nrow = 1)
 colnames(avg_dur) = c("All", "Control group", "JSIE",
 "Complier JSIE", "Non-complier JSIE", "HIE", "Complier HIE", "Non-complier HIE")
 avg_dur[1,1] = round(mean(wkpaid),2)
 avg_dur[1,2] = round(mean(wkpaid[W==0]),2)
 avg_dur[1,3] = round(mean(wkpaid[W==1]),2)
```

```
 avg_dur[1,6] = round(mean(wkpaid[W==2]),2)
 avg_dur[1,4] = round(mean(wkpaid[Z==1]),2)
 avg_dur[1,7] = round(mean(wkpaid[Z==2]),2)
 avg_dur[1,5] = round(mean(wkpaid[W==1&Z!=1]),2)
 avg_dur[1,8] = round(mean(wkpaid[W==2&Z!=2]),2)

 #this command gives the latex code for table 6
 print(xtable(avg_dur, type = "latex", digits = 2), file = "avg_dur.tex")

 #T-tests
 t.test(wkpaid[Z==0],wkpaid[Z==1])
 t.test(wkpaid[Z==0],wkpaid[Z==2])

 #censoring for  T >= 26
cutoff = 26
wkpaid = pmin(wkpaid, cutoff)
table(delta)

#defining the range of Z, W and duration
wkpaid.range = sort(unique(wkpaid))
Z.range = as.numeric(sort(unique(Z)))
W.range = as.numeric(sort(unique(W)))


#Estimation of S---------------------------------

#Computation of the terms of Kaplan-Meier product (see paper for notation)
Nzwt = array(0, dim = c(length(Z.range),length(W.range),length(wkpaid.range)))
Yzwt = array(0, dim = c(length(Z.range),length(W.range),length(wkpaid.range)))
Yzw=array(0, dim = c(length(Z.range),length(W.range)))
Yw=array(0, dim = length(W.range))
dNzwt = array(0, dim = c(length(Z.range),length(W.range),length(wkpaid.range)))

for (z in Z.range){
  for (w in W.range){
    for (wkp in wkpaid.range){
      Nzwt[which(Z.range ==z),which(W.range ==w),which(wkpaid.range ==wkp)] =
        sum(Z==z & W == w & wkpaid <= wkp & delta == 1)
      Yzwt[which(Z.range ==z),which(W.range ==w),which(wkpaid.range ==wkp)] =
        sum(Z==z & W == w & wkpaid >= wkp )
    }
  }
}

for(w in W.range){
  Yw[which(W.range ==w)] = sum(W == w)
  for(z in Z.range){
    Yzw[which(Z.range ==z),which(W.range ==w)] = sum(Z == z & W == w)
  }
}
```

```
dNzwt[,,1] = Nzwt[,,1]
dNzwt[,,2:(dim(Nzwt)[3])] = Nzwt[,,2:dim(Nzwt)[3]] - Nzwt[,,1:(dim(Nzwt)[3]-1)]


#Kaplan-Meier product
S_KM = array(1, dim = c(length(unique(Z)),length(unique(W)),length(wkpaid.range)))
for (z in Z.range){
  for (w in W.range){
    for (k in 1:(length(wkpaid.range)-1)){
      #if(sum(Yzwt[i,j,1:k] == 0) ==0){
      S_KM[which(Z.range ==z),which(W.range ==w),k+1] =
        prod(1-dNzwt[which(Z.range ==z),which(W.range ==w),1:k]/
        Yzwt[which(Z.range ==z),which(W.range ==w),1:k])
      #}
    }
  }
}

S_KM[!is.finite(S_KM)] = 0
a = S_KM[,,1]
a[z_w==0] = 0
S_KM[,,1] = a

#Kaplan-meier curves
data_used = data.frame(wkpaid,Z,W,delta)
par(mfrow=c(2,3))
for(z in 1:length(Z.range)){
  for (w in 1:length(W.range))
  {
    if (S_KM[z,w,1]==1)
    {
      selected_data = data_used[(data_used$Z == (z-1)) & (data_used$W == (w-1)),]
      kaplan_meier = survfit(Surv(selected_data$wkpaid, selected_data$delta) ~ 1,
      type = "kaplan-meier")
      plot(kaplan_meier,conf.int = F, las = 1, col = "blue", xlab = "weeks",
      ylab = "probability", main = paste("Z = ", Z.range[z], ", W = ", W.range[w]))
    }
  }
}


#Computation of smoothed S(t,z conditionally to w)
S_t_z = function(t,z,w){
  #Compute a smoothed version of S using an Epanechnikov kernel
  xs = 0:cutoff
  ys = S_KM[which(Z.range == z),which(W.range == w),]
  K = function(x){
    ifelse(abs(x)<=1,3/4*(1-(x)^2),0)
  }
  Kh = function(x,h = 2){return(K(x/h)/h)}
  St = sum(sapply(t-xs,Kh) * ys)/sum(sapply(t-xs,Kh))
```

```
  prob = St* (Yzw[which(Z.range == z),
                  which(W.range == w)]/Yw[which(W.range == w)])
  return(prob)
}


#curves of smoothed kaplan meier functions
par(mfrow=c(3,3))
for(z in 1:length(Z.range)){
  for (w in 1:length(W.range))
  {
      xs = seq(from=0,to=26,by=0.01)
      ys = vector(length = length(xs))
      for (i in 1:length(xs))
      {
        ys[i] = S_t_z(xs[i],z-1,w-1)
      }
      plot(xs,ys, col = "blue", xlab = "weeks", ylab = "probability", type = 'l',
           main = paste("Z = ", Z.range[z], ", W = ", W.range[w]))
  }
}

#Minimization of the norm of A ------------------------------------
A = function(theta, S){
  return(function(u){
    if(u<0 ){stop("invalid u")}
    res = array(0,dim = length(W.range))
    for(w in W.range){
      for(z in Z.range){
        res[which(W.range == w)] = res[which(W.range == w)] +
          S(theta[which(Z.range == z)],z = z,w = w)
      }
      res[which(W.range == w)] = res[which(W.range == w)] - exp(-u)
    }
    return(res)
  })
}

norm = function(x){sum(x^2)}

obj_function = function(theta, u){
  return(norm(A(theta,S_t_z)(u)))
}

#number of iterations
attempts = 20

#setting range of u (quantile = 1 - e^-u)
us = seq(0.01, 1, by = 0.01)
inputs = rep(us,attempts)
```

```
#making processing faster
cores=detectCores()
cl <- makeCluster(cores-1) #not to overload your computer
registerDoParallel(cl)


#calculating results
results <- foreach(i=1:length(inputs), .combine=rbind, .packages = packages)
%dopar% {
  u = inputs[i]
  est = optim(fn = obj_function, par = runif(length(Z.range), min = 0, max = cutoff),
    u=u, method = "L-BFGS-B", lower = c(0,0,0), upper = c(cutoff,cutoff,cutoff))

  return(c(u,est$par,est$value))
}


stopCluster(cl)
results = as.data.frame(results)
names(results) = c("u", sapply(Z.range, function(z){paste("theta",z, sep = "")} ),
"value")

#choosing optimal estimator
est.theta = as.data.frame(matrix(NA, nrow = length(us), ncol = ncol(results)))
names(est.theta) = names(results)
for(i in 1:length(us)){
 current = results[results$u == us[i],]
 est.theta[i,] = current[which.min(current$value),]
}

#Computation and plots of QTEs-----------------------

#Compuation of QTE
QTE_HIE = est.theta[,"theta2"]-est.theta[,"theta0"]
QTE_JSIE = est.theta[,"theta1"]-est.theta[,"theta0"]


#plot of the value of the objective function
par(mfrow = c(1,1))
plot(us, est.theta$value, main = "Value of the objective function :
full year duration", type = "l", col = "blue", ylab="", xlab = "u")

#plots of the QTE
plot(1-exp(-us),QTE_HIE, main = "QTE HIE", type = "l", col = "blue",
ylab = "weeks", xlab = "quantile",xlim = c(0,0.5))

plot(1-exp(-us),QTE_JSIE, main = "QTE JSIE", type = "l", col = "blue",
ylab = "weeks", xlab = "quantile",xlim = c(0,0.5))
```

B.2 Bootstrapping procedure

```
    #Bootstrapping-------------------------

#constructing estimator function (same as in previous section)
main = function(dataset){
  dataset = as.data.frame(dataset)
  dataset$lagree[is.na(dataset$lagree)] = 0
  wkpaid= as.numeric(dataset$wkspdbye) #spells
  Z = vector(mode = "numeric", length = nrow(dataset)) #Treatment variable
  W = vector(mode = "numeric", length = nrow(dataset)) #Instrumental variable
  delta = vector(mode = "numeric", length = nrow(dataset)) #Censoring indicator
  for (i in 1:nrow(dataset))
  {
    if ((dataset$hie[i]==1)&(dataset$lagree[i]==1))
    {
      Z[i] = 2
    }
    if ((dataset$jsie[i]==1)&(dataset$lagree[i]==1))
    {
      Z[i] = 1
    }
    if ((dataset$hie[i]==1))
    {
      W[i] = 2
    }
    if ((dataset$jsie[i]==1))
    {
      W[i] = 1
    }
    if (wkpaid[i]<=26)
    {
      delta[i] = 1
    }
  }
  z_w = table(Z,W)
  cutoff = 26
  wkpaid = pmin(wkpaid, cutoff)
  delta = as.factor(delta)
  wkpaid.range = sort(unique(wkpaid))
  Z.range = as.numeric(sort(unique(Z)))
  W.range = as.numeric(sort(unique(W)))
  Nzwt = array(0, dim = c(length(Z.range),length(W.range),length(wkpaid.range)))
  Yzwt = array(0, dim = c(length(Z.range),length(W.range),length(wkpaid.range)))
  Yzw=array(0, dim = c(length(Z.range),length(W.range)))
  Yw=array(0, dim = length(W.range))
  dNzwt = array(0, dim = c(length(Z.range),length(W.range),length(wkpaid.range)))
  for (z in Z.range){
    for (w in W.range){
      for (wkp in wkpaid.range){
```

```
      Nzwt[which(Z.range ==z),which(W.range ==w),which(wkpaid.range ==wkp)] =
        sum(Z==z & W == w & wkpaid <= wkp & delta == 1)
      Yzwt[which(Z.range ==z),which(W.range ==w),which(wkpaid.range ==wkp)] =
        sum(Z==z & W == w & wkpaid >= wkp )
    }
  }
}

for(w in W.range){
  Yw[which(W.range ==w)] = sum(W == w)
  for(z in Z.range){
    Yzw[which(Z.range ==z),which(W.range ==w)] = sum(Z == z & W == w)
  }
}
dNzwt[,,1] = Nzwt[,,1]
dNzwt[,,2:(dim(Nzwt)[3])] = Nzwt[,,2:dim(Nzwt)[3]] - Nzwt[,,1:(dim(Nzwt)[3]-1)]
#Kaplan-Meier product
S_KM = array(1, dim = c(length(unique(Z)),length(unique(W)),length(wkpaid.range)))
for (z in Z.range){
  for (w in W.range){
    for (k in 1:(length(wkpaid.range)-1)){
      #if(sum(Yzwt[i,j,1:k] == 0) ==0){
      S_KM[which(Z.range ==z),which(W.range ==w),k+1] =
        prod(1-dNzwt[which(Z.range ==z),which(W.range ==w),1:k]
        /Yzwt[which(Z.range ==z),which(W.range ==w),1:k])
      #}
    }
  }
}
S_KM[!is.finite(S_KM)] = 0
a = S_KM[,,1]
a[z_w==0] = 0
S_KM[,,1] = a
#Computation of S(t,z conditionally to w)
S_t_z = function(t,z,w){
  #Compute a smoothed version of S using a kernel
  xs = 0:cutoff
  ys = S_KM[which(Z.range == z),which(W.range == w),]
  K = function(x){
    ifelse(abs(x)<=1,3/4*(1-(x)^2),0)
  }
  Kh = function(x,h = 2){return(K(x/h)/h)}
  St = sum(sapply(t-xs,Kh) * ys)/sum(sapply(t-xs,Kh))
  prob = St* (Yzw[which(Z.range == z),
                  which(W.range == w)]/Yw[which(W.range == w)])
  return(prob)
}
A = function(theta, S){
  return(function(u){
    if(u<0 ){stop("invalid u")}
    res = array(0,dim = length(W.range))
```

```
      for(w in W.range){
        for(z in Z.range){
          res[which(W.range == w)] = res[which(W.range == w)] +
            S(theta[which(Z.range == z)],z = z,w = w)
        }
        res[which(W.range == w)] = res[which(W.range == w)] - exp(-u)
      }
      return(res)
    })
  }

  norm = function(x){sum(x^2)}

  obj_function = function(theta, u){
    return(norm(A(theta,S_t_z)(u)))
  }

  attempts = 20
  us = seq(0.01, 0.71, by = 0.01)
  inputs = rep(us,attempts)
  cores=detectCores()
  cl <- makeCluster(cores-1) #not to overload your computer
  registerDoParallel(cl)
  results <- foreach(i=1:length(inputs), .combine=rbind, .packages = packages)
  %dopar% {
    u = inputs[i]
    est = optim(fn = obj_function, par = runif(length(Z.range),min = 0,max = cutoff),
    u=u, method = "L-BFGS-B", lower = c(0,0,0), upper = c(cutoff,cutoff,cutoff))
    return(c(u,est$par,est$value))
  }

  stopCluster(cl)
  results = as.data.frame(results)
  names(results) = c("u", sapply(Z.range, function(z){paste("theta",z, sep = "")} ),
  "value")
  est.theta = as.data.frame(matrix(NA, nrow = length(us), ncol = ncol(results)))
  names(est.theta) = names(results)
  for(i in 1:length(us)){
    current = results[results$u == us[i],]
    est.theta[i,] = current[which.min(current$value),]
  }
  return(est.theta[,c("theta0","theta1","theta2")])
}

#number of bootstrap samples
number_dataset = 1000 #(code needs to run a long time if 1000 datasets are used)

bootstrapped_res = array(0, dim = c(71,3,number_dataset))
illinois = as.matrix(illinois)
us = seq(0.01, 0.71, by = 0.01)
```

```
#sampling
for (i in 1:number_dataset)
{
  data = matrix(nrow = nrow(illinois), ncol = ncol(illinois))
  colnames(data) = colnames(illinois)
  for (j in 1:nrow(illinois))
  {
    index = sample(1:nrow(illinois), 1)
    data[j,] = illinois[index,]
  }
  bootstrapped_res[,,i] = as.matrix(main(data))
}

#quantile treatment effects
QTE_HIE_b = bootstrapped_res[,3,] - bootstrapped_res[,1,]
QTE_JSIE_b = bootstrapped_res[,2,] - bootstrapped_res[,1,]

for (i in 1:71)
{
  QTE_HIE_b[i,] = sort(QTE_HIE_b[i,])
  QTE_JSIE_b[i,] = sort(QTE_JSIE_b[i,])
}

#setting lower (2.5 percentile) and upper (97.5 percentile) limits
index_lower = round(number_dataset*0.025,0)
index_upper = round(number_dataset*0.975,0)
CI_lower_hie = QTE_HIE_b[,index_lower]
CI_lower_jsie = QTE_JSIE_b[,index_lower]
CI_upper_hie = QTE_HIE_b[,index_upper]
CI_upper_jsie = QTE_JSIE_b[,index_upper]
par(mfrow = c(1,2))

#plot QTEs of JSIE treatment with confidence intervals
plot(1-exp(-us), QTE_JSIE[1:71], type = "l", col = "blue", main = "JSIE",
ylab = "weeks", xlab = "quantile", ylim = c(-6,0))
polygon(c((1-exp(-us)),rev(1-exp(-us))),c(CI_lower_jsie,rev(CI_upper_jsie)),
col = "grey75", border = FALSE)

lines(1-exp(-us),QTE_JSIE[1:71], type = "l", col = "blue")
lines(1-exp(-us),CI_lower_jsie, col = "red", lty = 2)
lines(1-exp(-us),CI_upper_jsie, col = "red", lty = 2)

#plot QTEs of HIE treatment with confidence intervals
plot(1-exp(-us), QTE_HIE[1:71], type = "l", col = "blue", main = "HIE",
ylab = "weeks", xlab = "quantile", ylim = c(-4,1))
polygon(c((1-exp(-us)),rev(1-exp(-us))),c(CI_lower_hie,rev(CI_upper_hie)),
col = "grey75", border = FALSE)

lines(1-exp(-us),QTE_HIE[1:71], type = "l", col = "blue")
lines(1-exp(-us),CI_lower_hie, col = "red", lty = 2)
lines(1-exp(-us),CI_upper_hie, col = "red", lty = 2)
```

B.3 Numerical example kernel regression

```
install.packages("kedd")
library("kedd")

#fictional dataset
example  = c(212, 184, 168, 45, 30, 10)

#calculating bandwidt
bandwidth = h.amise(example,kernel = "epanechnikov")
bandwidth

#kernel function
kernel = function(x,xi){
  epan_kern = 3/4 * (1 - ( (x-xi)/315)^2)
  return(epan_kern)
}

#kernel densities
d1 = kernel(60,230)
d2 = kernel(60,120)
d3 = kernel(60,100)
d4 = kernel(60,45)
d5 = kernel(60,35)
d6 = kernel(60,20)

#sum of kernel densities
w_total = sum(d1,d2,d3,d4,d5,d6)

#kernel weights
w1 = d1/w_total
w2 = d2/w_total
w3 = d3/w_total
w4 = d4/w_total
w5 = d5/w_total
w6 = d6/w_total


weights = c(w1,w2,w3,w4,w5,w6)
weights

#results
result = sum(weights %*% example)
result
```

Bibliography

Baiocchi, M., Cheng, J. and Small, D. S. (2014), 'Instrumental variable methods for causal inference', *Statistics in medicine* **33**(13), 2297–2340.

Beyhum, J., Florens, J.-P. and Van Keilegom, I. (2021), 'Nonparametric instrumental regression with right censored duration outcomes', *Journal of Business & Economic Statistics* pp. 1–12.

Bijwaard, G. E. (2007), 'Instrumental variable estimation of treatment effects for duration outcomes'.

Bijwaard, G. E. and Ridder, G. (2005), 'Correcting for selective compliance in a re-employment bonus experiment', *Journal of Econometrics* **125**(1-2), 77–111.

Byrd, R., Lu, P., Nocedal, J. and Zhu, C. (2003), 'A limited memory algorithm for bound constrained optimization', *SIAM Journal on Scientific Computing* **16**.

Card, D. and Krueger, A. B. (2000), 'Minimum wages and employment: A case study of the fast-food industry in new jersey and pennsylvania: Reply', *The American Economic Review* **90**(5), 1397–1420.

Chaisemartin, C. (2017), 'Tolerating defiance? local average treatment effects without monotonicity: Tolerating defiance?', *Quantitative Economics* **8**, 367–396.

Chernozhukov, V., Fernandez-Val, I. and Kowalski, A. (2011), 'Quantile regression with censoring and endogeneity', *Journal of Econometrics* **186**.

Chernozhukov, V. and Hansen, C. (2013), 'Quantile models with endogeneity', *Annu. Rev. Econ.* **5**(1), 57–81.

Chetty, R. (2008), Moral hazard vs. liquidity and optimal unemployment insurance, Working Paper 13967, National Bureau of Economic Research.

Clark, T. G., Bradburn, M. J., Love, S. B. and Altman, D. G. (2003), 'Survival analysis part i: basic concepts and first analyses', *British journal of cancer* **89**(2), 232–238.

Coate, D. and Grossman, M. (1988), 'Effects of alcoholic beverage prices and legal drinking ages on youth alcohol use', *The Journal of Law and Economics* **31**(1), 145–171.

Dabrowska, D. M. (1989), 'Uniform consistency of the kernel conditional kaplan-meier estimate', *The Annals of Statistics* **17**(3), 1157–1167.

Davino, C., Furno, M. and Vistocco, D. (2014), *Quantile regression: theory and applications*, Vol. 988, John Wiley & Sons.

Dye, S. (2020), 'Quantile regression', *Towards Data Science, February* **13**.

Jenkins, S. P. (2005), 'Survival analysis', *Unpublished manuscript, Institute for Social and Economic Research, University of Essex, Colchester, UK* **42**, 54–56.

Kaplan, E. L. and Meier, P. (1958), 'Nonparametric estimation from incomplete observations', *Journal of the American Statistical Association* **53**(282), 457–481.

Koenker, R. (2005), *Quantile Regression*, Econometric Society Monographs, Cambridge University Press.

Kolsrud, J., Landais, C., Nilsson, P. and Spinnewijn, J. (2018), 'The optimal timing of unemployment benefits: Theory and evidence from sweden', *American Economic Review* **108**(4-5), 985–1033.

Le Cook, B. and Manning, W. G. (2013), 'Thinking beyond the mean: a practical guide for using quantile regression methods for health services research', *Shanghai archives of psychiatry* **25**(1), 55.

Mahmoud, H. F. (2019), 'Parametric versus semi and nonparametric regression models', *arXiv preprint arXiv:1906.10221* .

Meyer, B. D. (1996), 'What have we learned from the illinois reemployment bonus experiment?', *Journal of labor Economics* **14**(1), 26–51.

Pramanik, N. (2019), 'Kernel regression — with example and code', *Retrieved April 9, 2022, from https://towardsdatascience.com/kernel-regression-made-easy-to-understand-86caf2d2b844* .

Sant'Anna, P. (2016), 'Program evaluation with right-censored data', *SSRN Electronic Journal* .

Sant'Anna, P. H. C. (2020), 'Nonparametric tests for treatment effect heterogeneity with duration outcomes', *Journal of Business & Economic Statistics* pp. 1–17.

Wei, B., Peng, L., Zhang, M. and Fine, J. (2021), 'Estimation of causal quantile effects with a binary instrumental variable and censored data', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **83**.

Woodbury, S. A. and Spiegelman, R. G. (1987), 'Bonuses to workers and employers to reduce unemployment: Randomized trials in illinois', *The American Economic Review* **77**(4), 513–530.

UNIVERSITÉ CATHOLIQUE DE LOUVAIN
Louvain School of Management
Place des Doyens, 1 bte L2.01.01, 1348 Louvain-la-Neuve, Belgique | **www.uclouvain.be/lsm**