Replication Study: The Duration of Unemployment

Josue Velasco

2024-03-07

I will replicate some results documented in:

Lalive, R. van Ours, J. and J. Zweimuller (2006), “How Changes in Financial Incentives Affect the Duration of Unemployment.” The Review of Economic Studies, 73, 4, 1009-1038.

1. Research objectives

This study focuses on a groundbreaking exploration: the intricate relationship between adjustments to financial support for unemployed individuals and the duration of their unemployment. The investigation meticulously dissects a pivotal period in Austria’s history – the late 1980s, a time marked by significant reforms to the nation’s unemployment insurance system. This research gains relevance in light of the widespread transformations to unemployment benefits that have swept across Europe in recent years.

Theoretical Framework: Optimizing Job Search Under Unemployment Benefits

The study is informed by a theoretical framework that sheds light on the relationship between unemployment benefit parameters and job search behavior. This framework centers on the concept of optimal search intensity, which considers the expected costs and benefits associated with being unemployed versus being employed. Factors such as the level of unemployment benefits, search costs incurred during job hunting, prevailing labor market conditions, the likelihood of finding a job, and the risk of exhausting benefits before securing employment all play a crucial role in this optimization problem.

According to the theory, optimal search intensity involves striking a balance between the marginal costs and benefits of searching for a job. Initially, when unemployment begins, search intensity tends to be low. This is because the probability of finding a job before benefits expire is relatively high. However, as the duration of unemployment lengthens, the risk of benefit exhaustion increases. This prompts unemployed individuals to increase their job search efforts, intensifying their search activities. Once benefits are exhausted, search intensity stabilizes as individuals confront a more predictable environment. In this phase, the rate at which they exit unemployment becomes primarily dependent on search costs, the level of any remaining unemployment assistance, and prevailing labor market conditions.

The theory further explores into the impact of changes in the earnings replacement rate (RR) on job search behavior. A higher RR translates to a lower cost of being unemployed, particularly at the outset. This initially leads to a reduction in search intensity. However, as benefit expiration draws closer, the potential loss of more generous benefits incentivizes individuals to intensify their job search efforts. This results in a surge in search intensity that even surpasses the level observed with lower RR. This phenomenon, termed the “entitlement effect,” underscores the heightened value of securing a new job due to the mitigated financial consequences of job loss associated with a higher RR.

Empirical Analysis: The Austrian Case Study

The authors leverage a natural experiment in Austria to isolate the causal effects of specific policy changes on unemployment durations. The experiment involved variations in unemployment benefit parameters for distinct groups of unemployed individuals:

- **Group 1:** Received a better deal on their benefits (increased RR).
- **Group 2:** Could get benefits for longer (extended PBD).
- **Group 3:** Got both the better deal and the longer benefits.
- **Group 4:** The control group – nothing changed for them with their benefits.

By meticulously comparing these groups' responses to the policy changes, the study sheds light on the intricate relationship between financial incentives and job search behavior.

Impact of RR and PBD on Unemployment Exit Hazard and Job Search Behavior

The excerpt explains how changes in the replacement rate (RR) and potential benefit duration (PBD) impact the unemployment exit hazard and job search behavior over the duration of unemployment. Here's a breakdown of the key points:

1. **Results:**
    - Group 1 (RR Increase): Individuals experiencing an isolated increase in the earnings replacement rate by about 15% showed a corresponding increase in unemployment duration.
    - Group 2 (PBD Extension): The impact of extending PBD from 30 to 39 weeks resulted in a relatively small increase in unemployment duration. However, extending PBD from 30 to 52 weeks led to a much larger increase in unemployment duration.
    - Combined Effects: The joint increase in RR and PBD had varying effects. For the extension from 30 to 39 weeks, the impact was slightly larger than the sum of individual changes. In contrast, the joint increase from 30 to 52 weeks had a significantly larger effect than the sum of individual changes.
    - Group 4: Did not experience relevant changes.
2. **Empirical Implications:**

In general, the study showed that increasing in RR and extensions in PBD are expected to result in longer unemployment durations. The strength of behavioral responses will vary depending on the specific changes in benefits and the stage of unemployment like the individual characteristics such as age, proximity to retirement, and eligibility for early retirement benefits. Additionally, labor market positions and institutional settings can also play a role.

Policy Implications

By incorporating this theoretical framework alongside their empirical analysis, the authors provide a comprehensive understanding of the complex dynamics between unemployment benefit parameters and job search behavior. This combined approach offers valuable insights for policymakers who are striving to design unemployment benefit programs that effectively balance the need for social support with the goal of facilitating re-employment and reducing unemployment durations.

The study highlights that changes in RR and PBD can significantly influence unemployment exit patterns. While extensions in PBD may not have a strong initial impact, they become more influential as unemployment progresses. The "entitlement effect" associated with higher RR can lead to surges in search intensity near benefit expiration. Additionally, the study suggests that the combined effect of increasing both RR and PBD can have a magnified disincentive on job search, particularly towards the end of the benefit period.

These findings underscore the importance of considering the relationship between benefit parameters and individual behavior when designing unemployment insurance systems. Policymakers can leverage this knowledge to create programs that provide adequate support while also encouraging active job search efforts, ultimately promoting faster re-employment and a healthier labor market.

2 Background

The authors seek to identify the causal effect of benefit duration on the willingness of individuals to accept jobs using a policy change that took place in Austria in 1989.

The policy affected various unemployed workers differently, as mentioned before: a first group experienced an increase in RR (replacement rate); a second group experienced an extension of PBD (potential benefit duration); a third group experienced both changes; and a fourth group experienced no change (the control group).

The potential benefit duration was increased, depending on age and experience: For workers younger than 40 and who had little previous work experience, the potential benefit duration remained unchanged. For workers with high levels of previous work experience, the duration has increased.

3. Load and prepare the data

The data are provided in the data set fi.dta. This file, which contains 225,821 unemployment spells, is quite large (150 MB) as it also contains the interaction terms used in the PH model estimation.

```
#load libraries to be used
library(foreign)
library(tidyverse)
```

```
## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──
## ✓ dplyr     1.1.4     ✓ readr     2.1.4
## ✓ forcats   1.0.0     ✓ stringr   1.5.1
## ✓ ggplot2   3.4.4     ✓ tibble    3.2.1
## ✓ lubridate 1.9.3     ✓ tidyr     1.3.0
## ✓ purrr     1.0.2
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## ✕ dplyr::filter() masks stats::filter()
## ✕ dplyr::lag()    masks stats::lag()
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to becom
e errors
```

```
library(survival)

data = read.dta('C:/Users/Josue/Documentos/Maestría/Aix-Marseille/Transition and duration model
s/fi.dta')
data <- data[,1:134] # get rid of some superfluous variables
data = as_tibble(data)

dim(data)
```

```
## [1] 225821    134
```

```
glimpse(data[,1:36])
```

```
## Rows: 225,821
## Columns: 36
## $ beginn   <dbl> 10412, 10169, 10624, 10343, 10654, 10101, 10652, 10338, 10705…
## $ ein_zus  <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0…
## $ sfrau    <dbl> 0, 1, 1, 0, 0, 0, 1, 1, 1, 0, 0, 1, 1, 1, 1, 1, 1, 0, 1, 1, 0…
## $ age      <dbl> 51.13758, 54.71595, 54.04791, 54.11636, 46.13005, 42.11909, 5…
## $ after    <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0…
## $ dur      <dbl> 0.1428571, 0.1428571, 0.1428571, 0.1428571, 0.1428571, 0.1428…
## $ nwage_pj <dbl> 11282.729, 8577.623, 6828.051, 10071.884, 6342.333, 6094.196,…
## $ e3_5     <dbl> 1, 1, 1, 0, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1…
## $ bdur     <dbl> 30, 30, 30, 20, 30, 30, 30, 20, 30, 30, 30, 30, 30, 30, 30, 3…
## $ y1988    <dbl> 1, 0, 0, 1, 0, 0, 0, 1, 0, 1, 1, 0, 0, 1, 1, 1, 0, 1, 0, 1, 1…
## $ y1989    <dbl> 0, 0, 1, 0, 1, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0…
## $ y1990    <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0…
## $ y1991    <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0…
## $ med_educ <dbl> 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0…
## $ hi_educ  <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0…
## $ lehre    <dbl> 1, 0, 0, 0, 1, 1, 0, 0, 1, 0, 1, 1, 0, 0, 0, 0, 1, 0, 0, 1, 0…
## $ married  <dbl> 1, 0, 1, 0, 0, 0, 1, 0, 0, 1, 0, 0, 1, 1, 1, 1, 1, 0, 1, 1, 1…
## $ single   <dbl> 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0…
## $ divorced <dbl> 0, 0, 0, 1, 1, 1, 0, 1, 1, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0…
## $ f_marr   <dbl> 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 0, 1, 1, 0…
## $ f_single <dbl> 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0…
## $ f_divor  <dbl> 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0…
## $ bc       <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 0, 1, 1, 1, 1, 0, 1, 0, 1, 1…
## $ pnon_10  <dbl> 0.2651971579, 0.1003559679, 0.1081599146, 0.3736308813, 0.033…
## $ q2       <dbl> 0, 0, 0, 1, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 1…
## $ q3       <dbl> 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0…
## $ q4       <dbl> 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0…
## $ seasonal <dbl> 1, 1, 1, 0, 0, 1, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0…
## $ manuf    <dbl> 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 1, 0, 1, 0, 0, 0, 0…
## $ ten72    <dbl> -3.47749186, -1.04384339, 0.23723496, -2.63423371, -2.5197806…
## $ type     <fct> PBD and RR, PBD and RR, PBD and RR, PBD and RR, PBD and RR, P…
## $ uncc     <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1…
## $ lwage    <dbl> 6.173206, 5.918063, 5.645761, 6.059680, 5.571968, 5.576242, 5…
## $ tr       <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1…
## $ t39      <dbl> 0, 0, 0, 0, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 0, 1, 0…
## $ t52      <dbl> 1, 1, 1, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 1…
```

Examining the groups defined for the study

```
table(data$type)
```

```
##
## PBD and RR        PBD         RR    control
##      21174      99404      32470      72773
```

Dealing with right censored data at 104 weeks:

$$t_u^{104} = min(t_u, 104)$$

To do this, we create a new column in the data frame.

```
#Compute average spells for durations bigger than 104 weeks (right-censored at 104 weeks)

data %>%
  mutate(dur104 = dur,
         dur104 = ifelse(dur104 > 104,104, dur104)) -> data
```

Description of variables

The variable $uncc$ refers to the exit indicator. While the variable $after$, which is a binary variable, refers to the period after the policy change.

The RDD design is as follows:

Previously (before August 1st, 1989):

- Unemployment benefits (PBD: Potential Benefit Duration) lasted 20 weeks for everyone.
- The RR (replacement rate) was around 41%.

Changes implemented after August 1st, 1989:

- RR increased to around 47%.
- The duration of unemployment benefits (PBD) became dependent on two factors:

  °Previous contributions (as it was before)

  °Age at the beginning of unemployment:

  - For individuals aged 40-49, the benefit duration increased to 39 weeks (if they had been employed within the past 10 years).
  - For individuals aged 50 and over, the benefit duration increased to a maximum of 52 weeks.

The variables and interaction terms are described as follows:

Variable Name	Description
dur	Duration of unemployment spell (weeks)
bdur	Potential benefit duration (weeks)
uncc	=1 if spell not censored
tr	=1 if replacement rate change
t39	=1 if PBD 30-39 change

Variable Name	Description
t52	=1 if PBD 30-52 change
t39_tr	t39 * tr
t52_tr	t52 * tr
tr_a0	tr * after0
t39_a0	t39 * after0
t52_a0	t52 * after0
t39tra0	t39 * tr * after0
t52tra0	t52 * tr * after0
after	=1 if spell starts after Aug 1, 1989
after0	= 1 if interval 0 after Aug 1, 1989

4. Difference-in-Differences

$$\Delta_{DD} = (\bar{Y}_A^{-T} - \bar{Y}_B^{-T}) - (\bar{Y}_A^{-C} - \bar{Y}_B^{-C})$$

$\bar{Y}_B^{-T}$ and $\bar{Y}_A^{-T}$ are the average duration of unemployment for the treated group before and after the implementation of the policy.

$\bar{Y}_A^{-C}$ and $\bar{Y}_B^{-C}$ refer to the average duration of unemployment for the control group (before and after the policy).

```
# Filter observations for Period after Aug 1, 1989)
period_after = data %>%
  filter(dur104 <= 104)%>% # focus on data less than or equal to 104 weeks
  filter(after==1)

#Filter observations for Period before Aug 1, 1989)
period_before = data %>%
  filter(dur104 <= 104)%>%
  filter(after==0)
```

```r
library(dplyr)
library(plotrix)

# Define a function to calculate mean, number of observations (n), and standard deviation (sd)
calculate_stats = function(data, groups, values) {
  result = data %>%
    group_by({{groups}}, after) %>%
    summarise(
      m = mean({{values}}),
      N = n(),
      se = std.error({{values}})
    ) %>%
    mutate(
      group = ifelse(after == 1, paste({{groups}}, "after", sep = " "), paste({{groups}}, "befor
e", sep = " "))
    ) %>%
    select(-after)
  return(result)
}

# Filter only 'PBD' and 'RR' from the 'type' column
period_before_filtered = period_before %>%
  filter(type %in% c('PBD', 'RR'))

period_after_filtered = period_after %>%
  filter(type %in% c('PBD', 'RR'))

# Calculate statistics for 'PBD' and 'RR' before and after
group_stats_before = calculate_stats(period_before_filtered, groups = type, values = dur104)
```

```
## `summarise()` has grouped output by 'type'. You can override using the
## `.groups` argument.
```

```r
group_stats_after = calculate_stats(period_after_filtered, groups = type, values = dur104)
```

```
## `summarise()` has grouped output by 'type'. You can override using the
## `.groups` argument.
```

```r
# Combine the results
combined_stats = bind_rows(group_stats_before, group_stats_after)

# Print the result
print(combined_stats)
```

```
## # A tibble: 4 × 5
## # Groups:   type [2]
##   type      m     N     se group
##   <fct> <dbl> <int>  <dbl> <chr>
## 1 PBD    15.8 48294 0.0757 PBD before
## 2 RR     17.1 17160 0.118  RR before
## 3 PBD    18.1 51110 0.0912 PBD after
## 4 RR     19.1 15310 0.152  RR after
```

By doing this, we get the means, standard error and number of observations for the groups RR and PBD.

We can replicate the same for all 4 groups (PDB, RR, PBD & RR and Control, both for before and after the implementation of the policy).

```
# Calculate statistics for 'PBD' and 'RR' before and after
group_stats_before = calculate_stats(period_before, groups = type, values = dur104)
group_stats_after = calculate_stats(period_after, groups = type, values = dur104)

# Combine the results
combined_stats = bind_rows(group_stats_before, group_stats_after)

# Print the result
print(combined_stats)
```

```
## # A tibble: 8 × 5
## # Groups:   type [4]
##   type           m     N     se group
##   <fct>      <dbl> <int>  <dbl> <chr>
## 1 PBD and RR  18.5 11992 0.162  PBD and RR before
## 2 PBD         15.8 48294 0.0757 PBD before
## 3 RR          17.1 17160 0.118  RR before
## 4 control     14.5 33815 0.0782 control before
## 5 PBD and RR  22.7  9182 0.233  PBD and RR after
## 6 PBD         18.1 51110 0.0912 PBD after
## 7 RR          19.1 15310 0.152  RR after
## 8 control     15.6 38958 0.0870 control after
```

Now, we can calculate the difference-in-difference for each group and get the following results:

```r
library(dplyr)
library(tidyr)
library(kableExtra)
library(knitr)

# Define a function to calculate mean, number of observations (n), and standard deviation (sd)
calculate_stats_long = function(data, values) {
  result = data %>%
    group_by(type, after) %>%
    summarise(
      m = mean({{values}}),
      N = n(),
      se = std.error({{values}})
    )
  return(result)
}

# Calculate statistics for all 'type' values before and after
group_stats_before_long = calculate_stats_long(period_before, values = dur104)
group_stats_after_long = calculate_stats_long(period_after, values = dur104)

# Combine the results
combined_stats = bind_rows(group_stats_before_long, group_stats_after_long)

# Pivot to long format
combined_stats_long = combined_stats %>%
  pivot_longer(cols = c(m, N, se), names_to = "name", values_to = "value")%>%pivot_wider(names_f
rom = after, values_from = value)

# Calculate differences and DiD
control_diff = combined_stats_long %>%
  filter(type == "control") %>%
  mutate(diff = `1` - `0`) %>%
  pull()

combined_stats_long = combined_stats_long %>%
  mutate(diff = ifelse(name == "m", round(`1` - `0`, 2), NA),
         DiD = ifelse(name == "m",
                      round(`1` - `0` - control_diff, 2),
                      NA))
# Print the result
print(combined_stats_long)
```

```
## # A tibble: 12 × 6
## # Groups:   type [4]
##    type       name         `0`        `1`  diff   DiD
##    <fct>      <chr>      <dbl>      <dbl> <dbl> <dbl>
##  1 PBD and RR m        18.5       22.7     4.25  3.08
##  2 PBD and RR N     11992       9182      NA    NA
##  3 PBD and RR se        0.162      0.233  NA    NA
##  4 PBD        m        15.8       18.1     2.25  1.08
##  5 PBD        N     48294      51110      NA    NA
##  6 PBD        se        0.0757     0.0912 NA    NA
##  7 RR         m        17.1       19.1     1.99  0.82
##  8 RR         N     17160      15310      NA    NA
##  9 RR         se        0.118      0.152  NA    NA
## 10 control    m        14.5       15.6     1.17  0
## 11 control    N     33815      38958      NA    NA
## 12 control    se        0.0782     0.0870 NA    NA
```

As reported by the original study, we can show too in this implementation that the duration of unemployment is bigger for the PBD & RR group, while the control group has the shortest one (both for before and after the implementation of the policy). The difference before and after the implementation show that unemployment duration increased in all four groups but this change is stronger within the groups unlike the control group (as showed by the DiD).

If we apply a regression, we get similar results:

```
# Fit separate linear regression models for each type
mod1 = lm(dur104 ~ type-1, data = period_after)
mod2 = lm(dur104 ~ type-1, data = period_before)
```

After the policy:

```
# Extract coefficients
mod1
```

```
##
## Call:
## lm(formula = dur104 ~ type - 1, data = period_after)
##
## Coefficients:
## typePBD and RR         typePBD          typeRR     typecontrol
##          22.74           18.08           19.10           15.63
```

Before the policy:

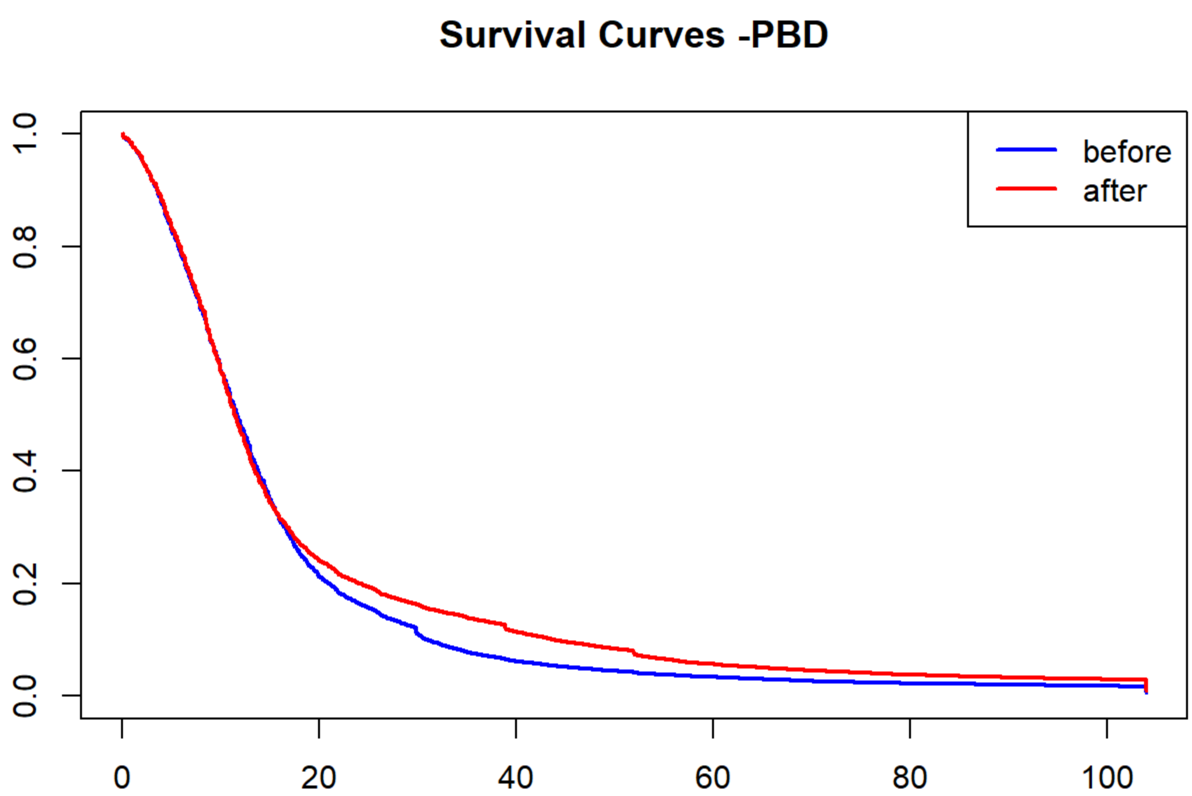
```
mod2
```

```
##
## Call:
## lm(formula = dur104 ~ type - 1, data = period_before)
##
## Coefficients:
## typePBD and RR         typePBD          typeRR     typecontrol
##          18.49           15.83           17.11           14.46
```

5. Survival Function

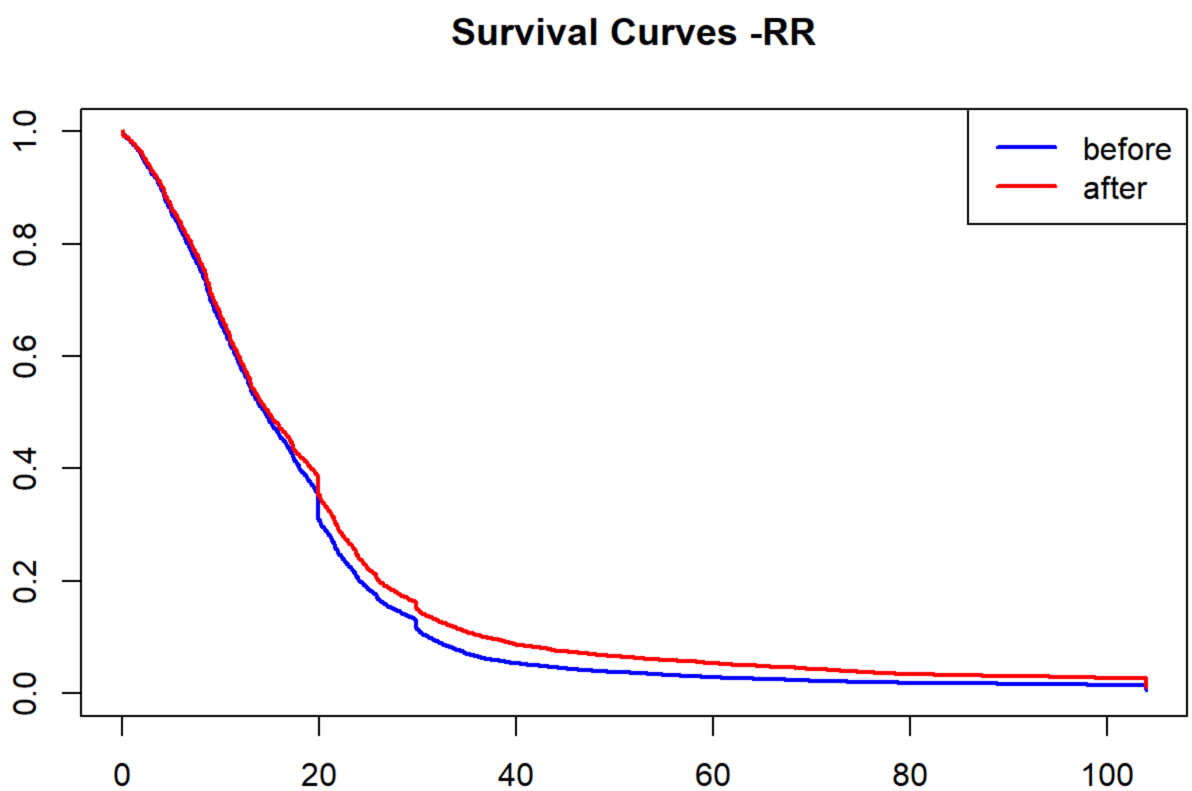
Now, I calculate and plot the Kaplan-Meier estimates (survival function); one for each group type.

```
period_pbd = data %>%
  filter(dur104 <= 104)%>%
  filter(type=="PBD")

fit_pbd = survfit(Surv(period_pbd$dur104, period_pbd$uncc)~period_pbd$after)

plot(fit_pbd, col = c("blue", "red"), lwd = 2, main = "Survival Curves -PBD")
legend("topright", legend = c("before","after"), col = c("blue", "red"), lwd = 2)
```



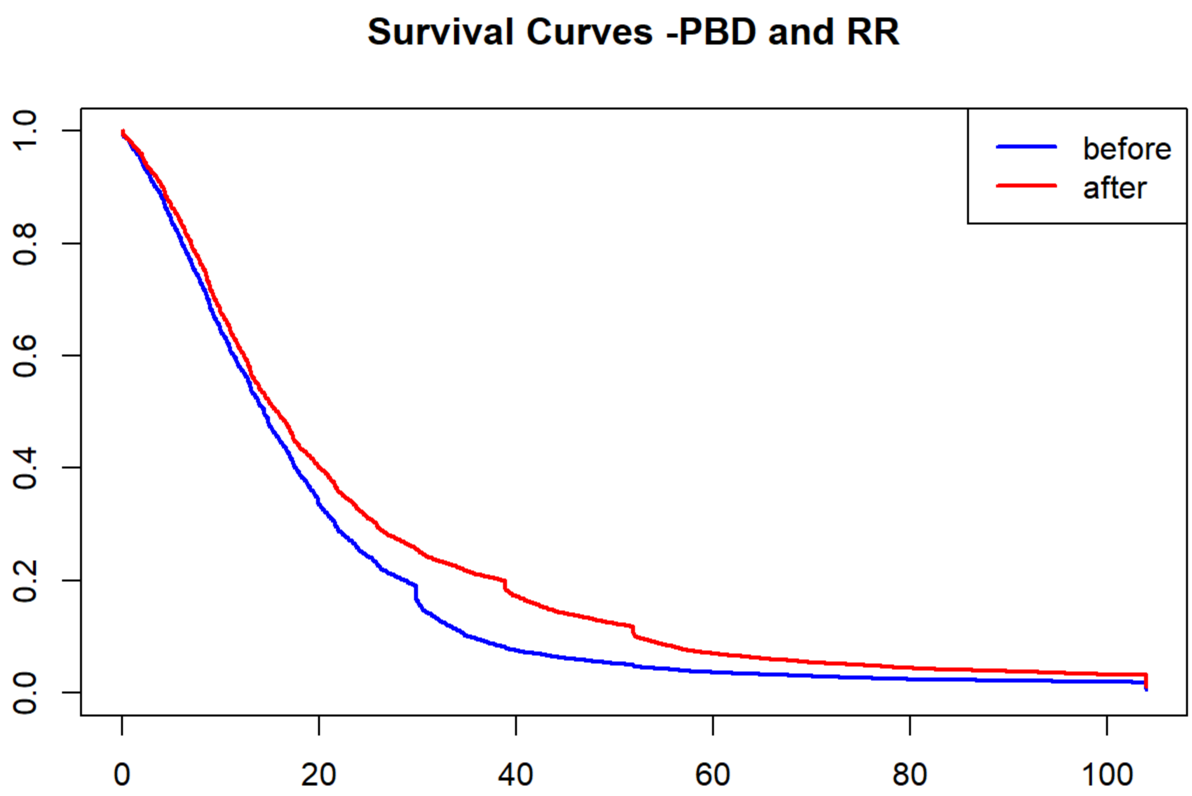
We get the same results as the original study in which for this group, after 15 weeks of unemployment, the survivor function after the policy change (red line), diverges from the survivor function before the policy (blue line), up to the 40th week.

```
period_rr = data %>%
  filter(dur104 <= 104)%>%
  filter(type=="RR")

fit_rr = survfit(Surv(period_rr$dur104, period_rr$uncc)~period_rr$after)

plot(fit_rr, col = c("blue", "red"), lwd = 2, main = "Survival Curves -RR")
legend("topright", legend = c("before","after"), col = c("blue", "red"), lwd = 2)
```



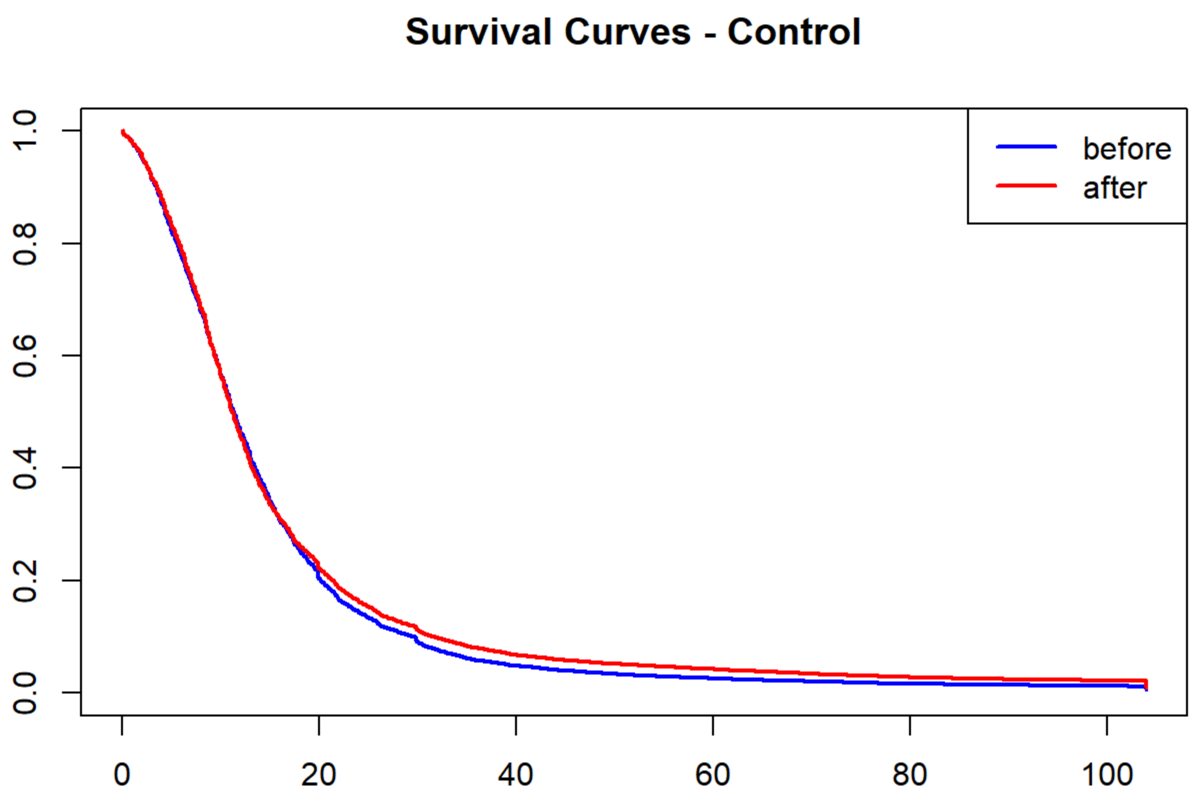
Unlike the previous graph, the RR group, there is a slight increase in the survivor function from the start of the spell for the red line (after the implementation of the policy), even after the week 30 in which the benefits have been exhausted.

```
period_pbd_rr = data %>%
  filter(dur104 <= 104)%>%
  filter(type=="PBD and RR")

fit_pbd_rr = survfit(Surv(period_pbd_rr$dur104, period_pbd_rr$uncc)~period_pbd_rr$after)

plot(fit_pbd_rr, col = c("blue", "red"), lwd = 2, main = "Survival Curves -PBD and RR")
legend("topright", legend = c("before","after"), col = c("blue", "red"), lwd = 2)
```



As for the PDB & RR group, there is a strong increase in the survivor function since the beginning of the unemployment spell, specially from the 15th week to the 40th, like for the PBD graph.

```
period_control = data %>%
  filter(dur104 <= 104)%>%
  filter(type=="control")

fit_control = survfit(Surv(period_control$dur104, period_control$uncc)~period_control$after)

plot(fit_control, col = c("blue", "red"), lwd = 2, main = "Survival Curves - Control")
legend("topright", legend = c("before","after"), col = c("blue", "red"), lwd = 2)
```



Finally, for the control group there is no much difference in the survivor functions.

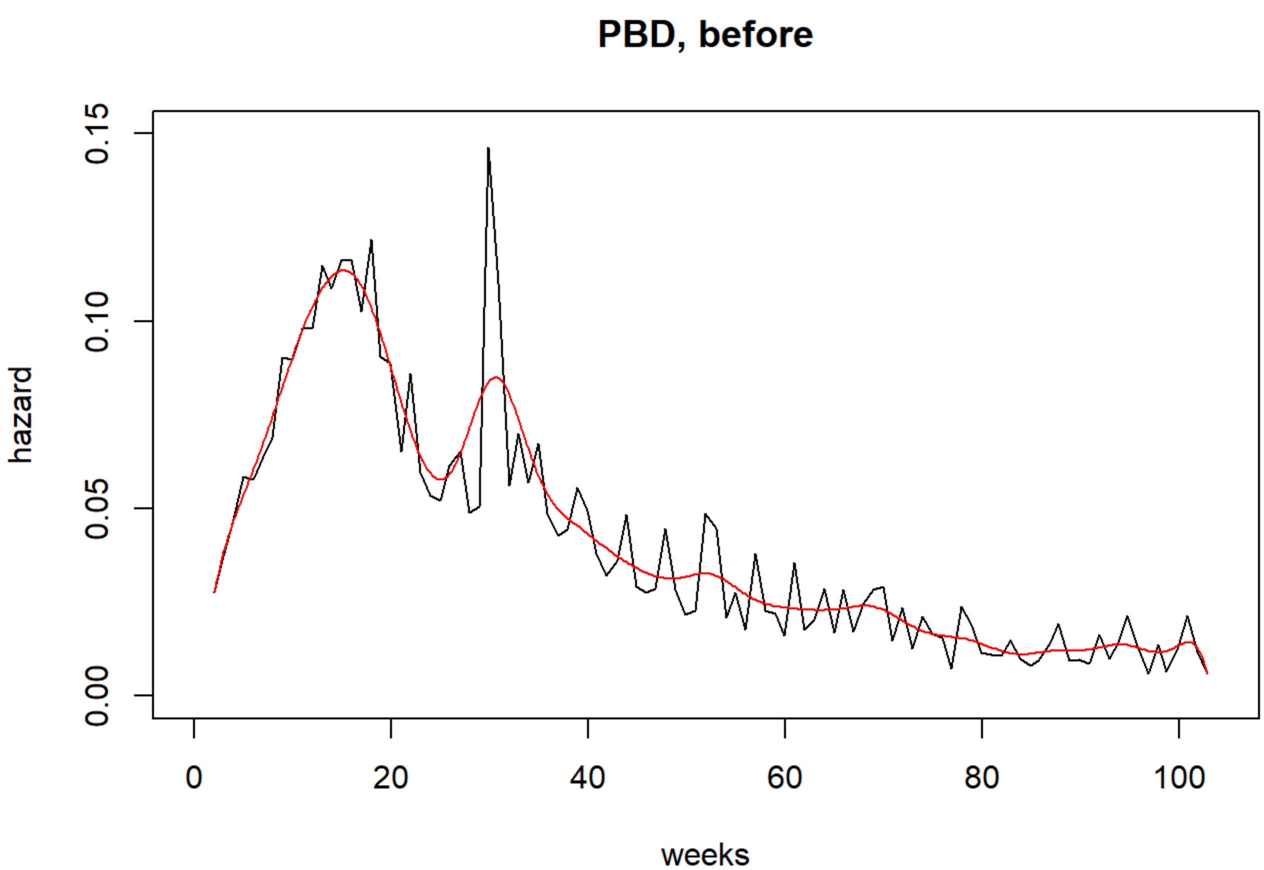
5.1 KM estimates of the unemployment exit hazard

Now, I can calculate the exit hazard rate with the previous estimates. Like in your implementation, I got the same graph for the PBD group before the implementation of the policy, successfully fitting the smooth graph by estimating the hazard as:

$$\hat{\lambda}(t) = \frac{\hat{f}(t)}{\hat{S}(t)}$$

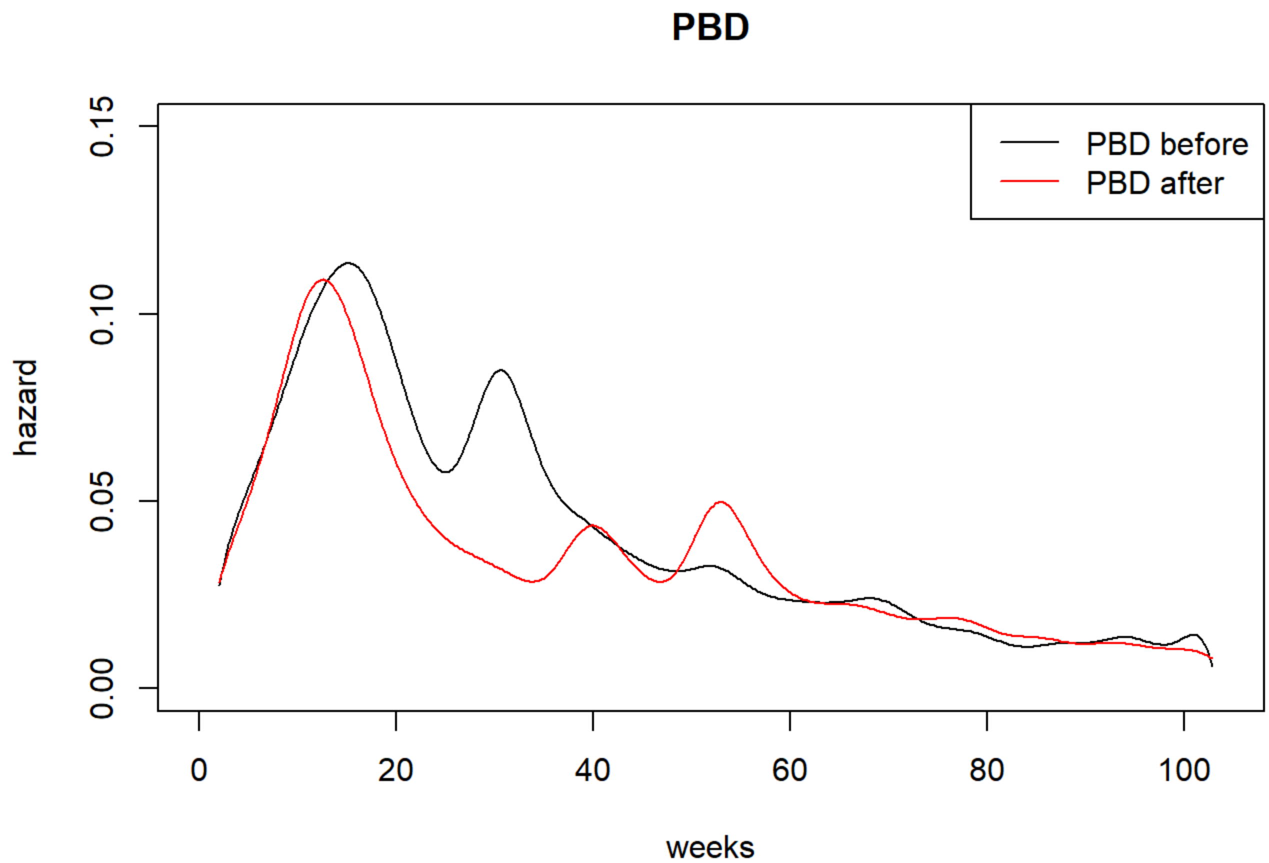
and using the locpoly function from the KernSmooth package.

```r
library(survival)
library(KernSmooth)

pbd_before =fit_pbd[1]
pbd_before = na.omit(pbd_before)
snew = data.frame(time=pbd_before$time, surv=pbd_before$surv)
snew$t_diff=c(NA, diff(snew$time,1))
snew$hz = c(NA, (-diff(snew$surv)/diff(snew$time,1))/snew$surv[-1])
snew$cum_hz=c(0,cumsum(snew$hz[-1]*snew$t_diff[-1]))

# n_weeks = sum(period_pbd$after == 0)
new.time=seq(from=0,to=103, by=1)

pbd_before$time = sort(pbd_before$time) #order time increasingly
i = findInterval(new.time, pbd_before$time)
snew_grid = data.frame(time=pbd_before$time[i], surv=pbd_before$surv[i])
snew_grid$t_diff = c(NA, diff(snew_grid$time, 1))
snew_grid$hz=c(NA,-(diff(snew_grid$surv,1)/diff(snew_grid$time,1))/snew_grid$surv[-1])

snew_grid$cum_hz = c(0,cumsum(snew_grid$hz[-1]*snew_grid$t_diff[-1]))

plot(c(0,104),c(0,0.15),xlab="weeks",ylab="hazard", main="PBD, before",type="n")
lines(snew_grid$time,snew_grid$hz)
h2=locpoly(snew_grid$time[-1],snew_grid$hz[-1],degree=3,bandwidth = 3)
lines(h2$x, h2$y, col="red")
```



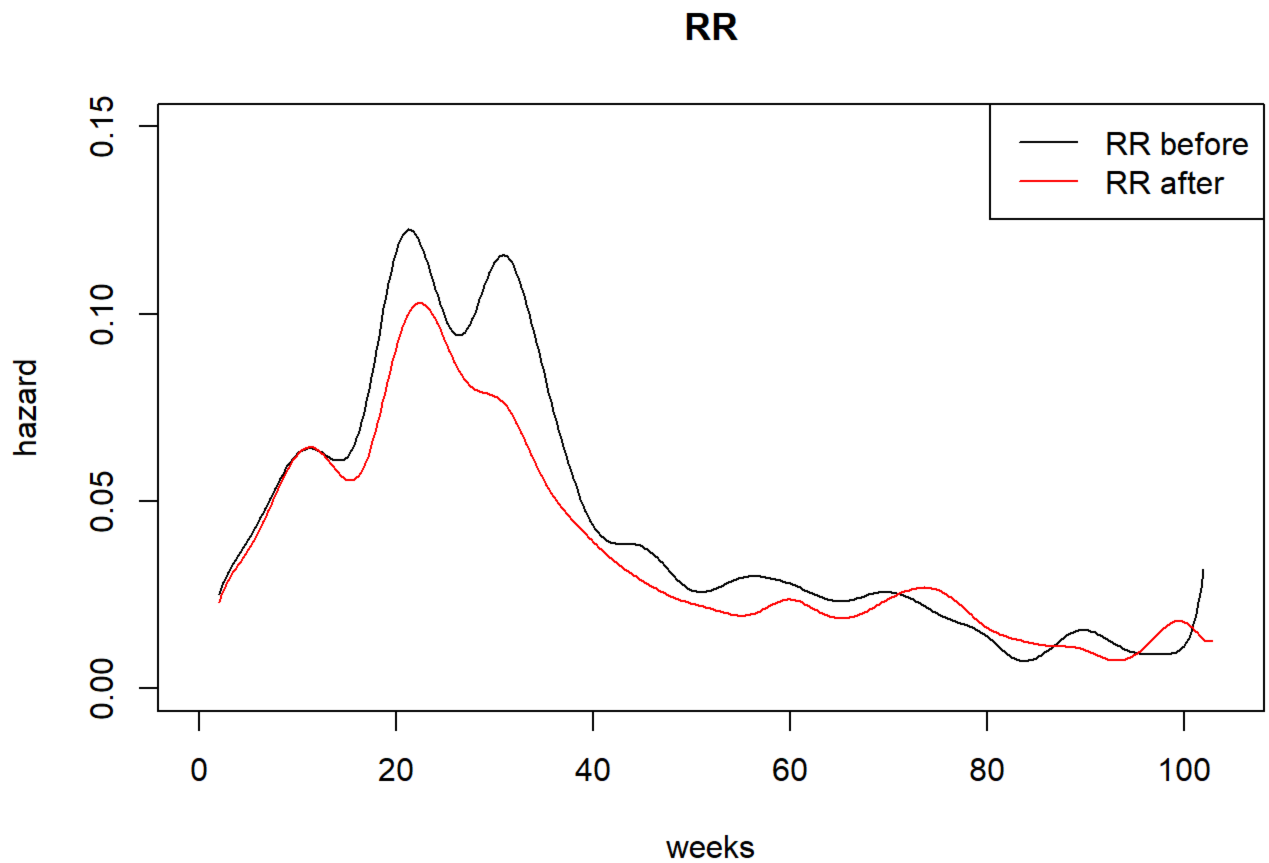
Then, I replicate this process for all the groups and plotting them together (I had to extract the survivor functions before and after the policy and get the hazards for each, I don’t know if there is a simpler way but I applied this for my implementation and worked well).

```
library(survival)
library(KernSmooth)

pbd_after =fit_pbd[2]
pbd_after = na.omit(pbd_after)
# snew_af = data.frame(time=pbd_after$time, surv=pbd_after$surv)
# snew_af$t_diff=c(NA, diff(snew_af$time,1))
# snew_af$hz = c(NA, (-diff(snew_af$surv)/diff(snew_af$time,1))/snew_af$surv[-1])

# n_weeks = sum(period_pbd$after == 0)
new.time=seq(from=0,to=103, by=1)

pbd_after$time = sort(pbd_after$time) #order time increasingly
i = findInterval(new.time, pbd_after$time)
snew_grid_af = data.frame(time=pbd_after$time[i], surv=pbd_after$surv[i])
snew_grid_af$t_diff = c(NA, diff(snew_grid_af$time, 1))
snew_grid_af$hz=c(NA,-(diff(snew_grid_af$surv,1)/diff(snew_grid_af$time,1))/snew_grid_af$surv[-
1])

# plot(c(0,104),c(0,0.15),xlab="weeks",ylab="hazard", main="PBD",type="n")
# # lines(snew_grid_af$time,snew_grid_af$hz)
h2_af=locpoly(snew_grid_af$time[-1],snew_grid_af$hz[-1],degree=3,bandwidth = 3)
# lines(h2_af$x, h2_af$y)
plot(c(0,104),c(0,0.15),xlab="weeks",ylab="hazard", main="PBD",type="n")
lines(h2$x, h2$y, col="black")
lines(h2_af$x, h2_af$y, col="red")
legend("topright", legend = c("PBD before", "PBD after"), col = c("black", "red"), lty = 1, cex=
1)
```



I got the same results as in the study in which the exit rate is higher (and sooner) before the implementation of the policy (black line), which are before the exhaustion of benefits (the spikes for the red line are also before the exhaustion of benefits).

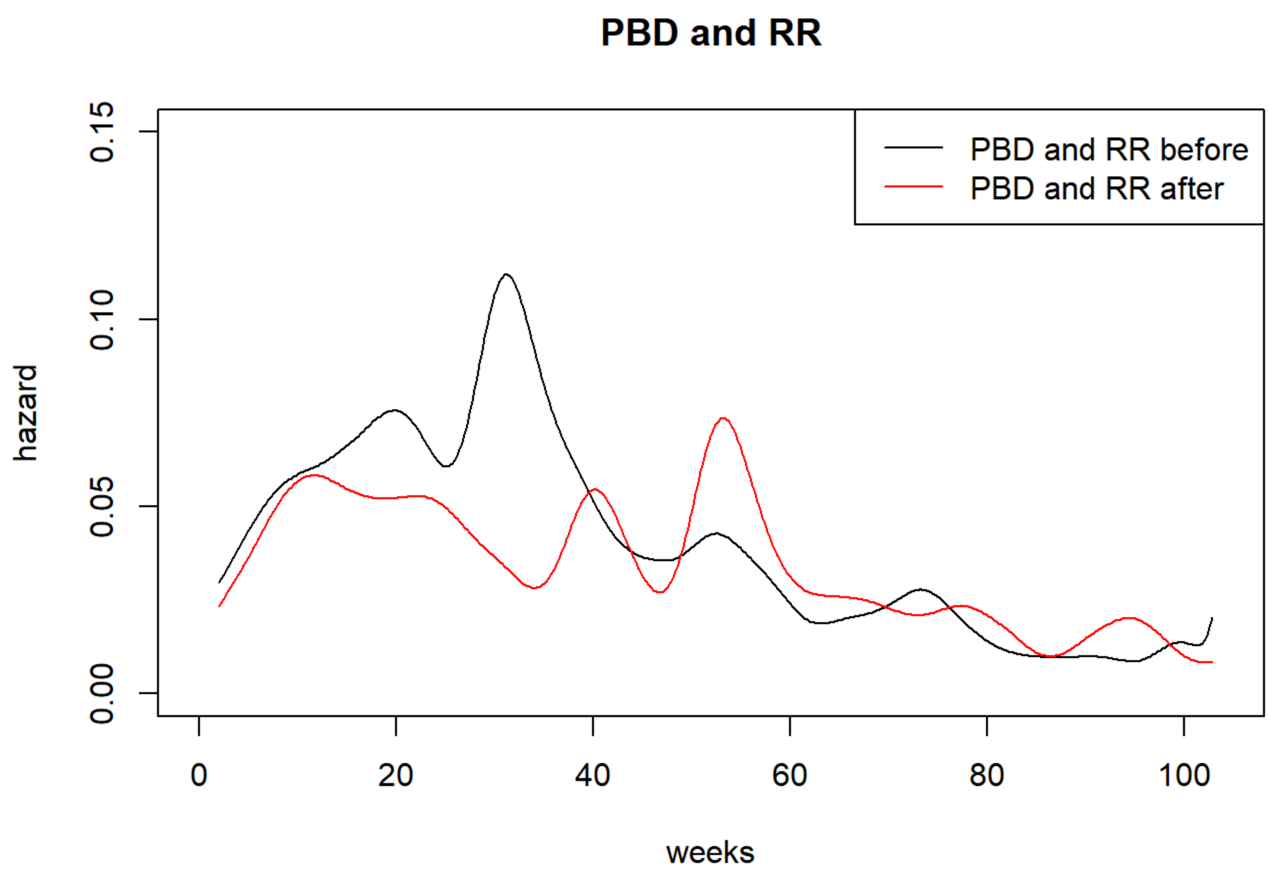
```
# separate KM estimates before and after the policy
rr_before = fit_rr[1]
rr_after =fit_rr[2]

# omit na values
rr_before = na.omit(rr_before)
rr_after = na.omit(rr_after)

# create date points
new.time=seq(from=0,to=103, by=1)

#order time increasingly
rr_before$time = sort(rr_before$time)
rr_after$time = sort(rr_after$time)

#find interval
i = findInterval(new.time, rr_before$time)
j = findInterval(new.time, rr_after$time)

#create data frames for before and after the implementation of the policies
snew_grid_rr_before = data.frame(time_before=rr_before$time[i], surv_before=rr_before$surv[i])
snew_grid_rr_after = data.frame(time_after=rr_after$time[j], surv_after=rr_after$surv[j])

#add columns for difference in periods for each dataframe
snew_grid_rr_before$t_diff_before = c(NA, diff(snew_grid_rr_before$time_before, 1))
snew_grid_rr_after$t_diff_after = c(NA, diff(snew_grid_rr_after$time_after, 1))

# calculate hazard rates for each dataframe
snew_grid_rr_before$hz_before=c(NA,-(diff(snew_grid_rr_before$surv_before,1)/diff(snew_grid_rr_b
efore$time_before,1))/snew_grid_rr_before$surv_before[-1])
snew_grid_rr_after$hz_after=c(NA,-(diff(snew_grid_rr_after$surv_after,1)/diff(snew_grid_rr_after
$time_after,1))/snew_grid_rr_after$surv_after[-1])

#omit NaN values
snew_grid_rr_before$hz_before =  replace(snew_grid_rr_before$hz_before, is.nan(snew_grid_rr_befo
re$hz_before), 0)

#plot both dataframes into a single one
plot(c(0,104),c(0,0.15),xlab="weeks",ylab="hazard", main="RR",type="n")

# smooth the estimator
h2_rr_before=locpoly(snew_grid_rr_before$time_before[-1],snew_grid_rr_before$hz_before[-1],degre
e=3,bandwidth = 3)
h2_rr_after=locpoly(snew_grid_rr_after$time_after[-1],snew_grid_rr_after$hz_after[-1],degree=3,b
andwidth = 3)

# add the lines for each graph
lines(h2_rr_before$x, h2_rr_before$y, col = c("black"))
lines(h2_rr_after$x, h2_rr_after$y, col=c("red"))

legend("topright", legend = c("RR before", "RR after"), col = c("black", "red"), lty = 1, cex=1)
```



The same occurs with the RR group in which there is a higher hazard for the old system.

```
# separate KM estimates before and after the policy
pbd_rr_before = fit_pbd_rr[1]
pbd_rr_after = fit_pbd_rr[2]

# omit na values
pbd_rr_before = na.omit(pbd_rr_before)
pbd_rr_after = na.omit(pbd_rr_after)

# create date points
new.time = seq(from = 0, to = 103, by = 1)

# order time increasingly
pbd_rr_before$time = sort(pbd_rr_before$time)
pbd_rr_after$time = sort(pbd_rr_after$time)

# find interval
i = findInterval(new.time, pbd_rr_before$time)
j = findInterval(new.time, pbd_rr_after$time)

# create data frames for before and after the implementation of the policies
snew_grid_pbd_rr_before = data.frame(time_before = pbd_rr_before$time[i], surv_before = pbd_rr_b
efore$surv[i])
snew_grid_pbd_rr_after = data.frame(time_after = pbd_rr_after$time[j], surv_after = pbd_rr_after
$surv[j])

# add columns for difference in periods for each dataframe
snew_grid_pbd_rr_before$t_diff_before = c(NA, diff(snew_grid_pbd_rr_before$time_before, 1))
snew_grid_pbd_rr_after$t_diff_after = c(NA, diff(snew_grid_pbd_rr_after$time_after, 1))

# calculate hazard rates for each dataframe
snew_grid_pbd_rr_before$hz_before = c(NA, -(diff(snew_grid_pbd_rr_before$surv_before, 1) / diff
(snew_grid_pbd_rr_before$time_before, 1)) / snew_grid_pbd_rr_before$surv_before[-1])
snew_grid_pbd_rr_after$hz_after = c(NA, -(diff(snew_grid_pbd_rr_after$surv_after, 1) / diff(snew
_grid_pbd_rr_after$time_after, 1)) / snew_grid_pbd_rr_after$surv_after[-1])

# omit NaN values
snew_grid_pbd_rr_before$hz_before = replace(snew_grid_pbd_rr_before$hz_before, is.nan(snew_grid_
pbd_rr_before$hz_before), 0)

# plot both dataframes into a single one
plot(c(0, 104), c(0, 0.15), xlab = "weeks", ylab = "hazard", main = "PBD and RR", type = "n")

# smooth the estimator
h2_pbd_rr_before = locpoly(snew_grid_pbd_rr_before$time_before[-1], snew_grid_pbd_rr_before$hz_b
efore[-1], degree = 3, bandwidth = 3)
h2_pbd_rr_after = locpoly(snew_grid_pbd_rr_after$time_after[-1], snew_grid_pbd_rr_after$hz_after
[-1], degree = 3, bandwidth = 3)

# add the lines for each graph
lines(h2_pbd_rr_before$x, h2_pbd_rr_before$y, col = c("black"))
lines(h2_pbd_rr_after$x, h2_pbd_rr_after$y, col = c("red"))
```

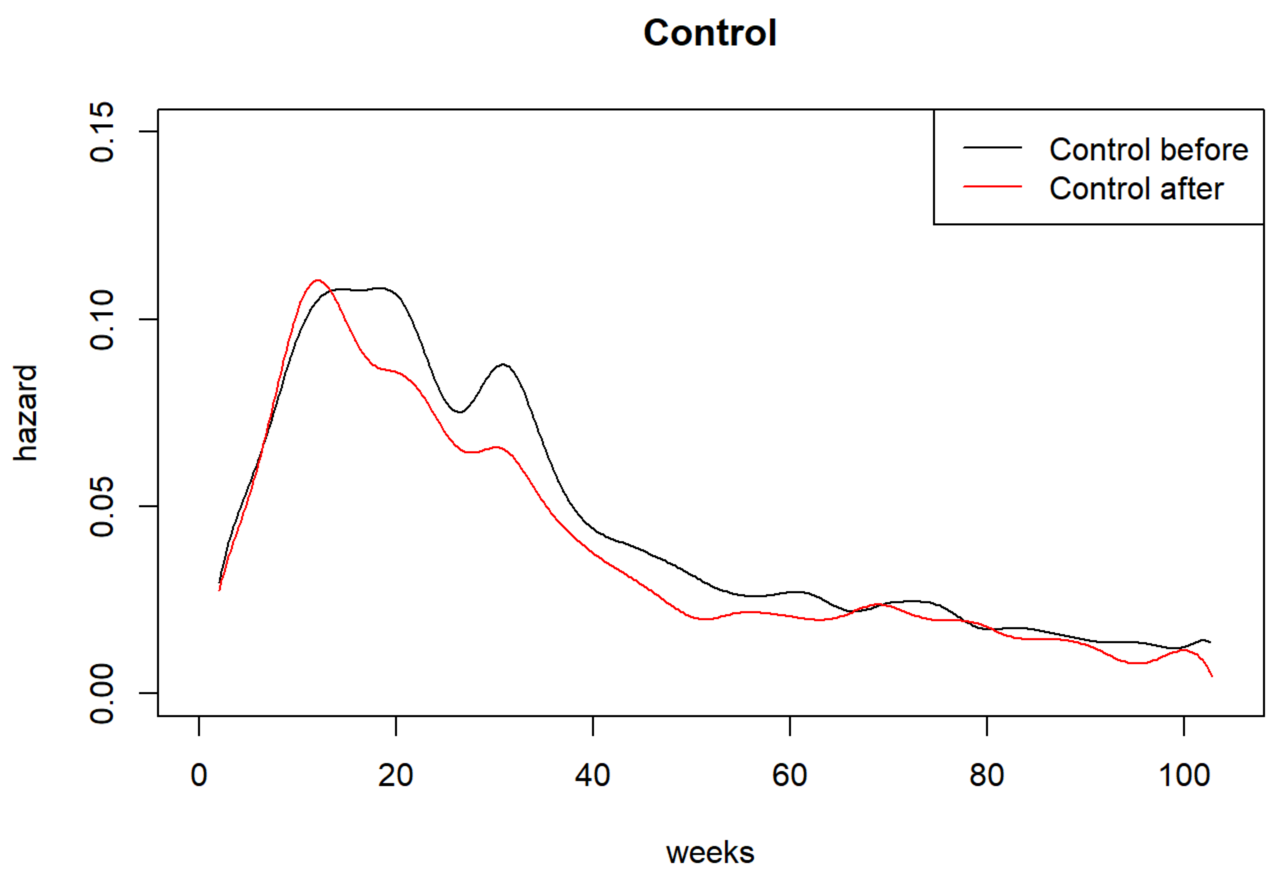
```
legend("topright", legend = c("PBD and RR before", "PBD and RR after"), col = c("black", "red"),
lty = 1, cex=1)
```



We can see this effect to in the combined effects of groups where there is a higher exit rate for the old system, specially, before the exhaustion of benefits.

```
# separate KM estimates before and after the policy
control_before = fit_control[1]
control_after = fit_control[2]

# omit na values
control_before = na.omit(control_before)
control_after = na.omit(control_after)

# create date points
new.time = seq(from = 0, to = 103, by = 1)

# order time increasingly
control_before$time = sort(control_before$time)
control_after$time = sort(control_after$time)

# find interval
i = findInterval(new.time, control_before$time)
j = findInterval(new.time, control_after$time)

# create data frames for before and after the implementation of the policies
snew_grid_control_before = data.frame(time_before = control_before$time[i], surv_before = contro
l_before$surv[i])
snew_grid_control_after = data.frame(time_after = control_after$time[j], surv_after = control_af
ter$surv[j])

# add columns for difference in periods for each dataframe
snew_grid_control_before$t_diff_before = c(NA, diff(snew_grid_control_before$time_before, 1))
snew_grid_control_after$t_diff_after = c(NA, diff(snew_grid_control_after$time_after, 1))

# calculate hazard rates for each dataframe
snew_grid_control_before$hz_before = c(NA, -(diff(snew_grid_control_before$surv_before, 1) / dif
f(snew_grid_control_before$time_before, 1)) / snew_grid_control_before$surv_before[-1])
snew_grid_control_after$hz_after = c(NA, -(diff(snew_grid_control_after$surv_after, 1) / diff(sn
ew_grid_control_after$time_after, 1)) / snew_grid_control_after$surv_after[-1])

# omit NaN values
snew_grid_control_before$hz_before = replace(snew_grid_control_before$hz_before, is.nan(snew_gri
d_control_before$hz_before), 0)

# plot both dataframes into a single one
plot(c(0, 104), c(0, 0.15), xlab = "weeks", ylab = "hazard", main = "Control", type = "n")

# smooth the estimator
h2_control_before = locpoly(snew_grid_control_before$time_before[-1], snew_grid_control_before$h
z_before[-1], degree = 3, bandwidth = 3)
h2_control_after = locpoly(snew_grid_control_after$time_after[-1], snew_grid_control_after$hz_af
ter[-1], degree = 3, bandwidth = 3)

# add the lines for each graph
lines(h2_control_before$x, h2_control_before$y, col = c("black"))
lines(h2_control_after$x, h2_control_after$y, col = c("red"))
```

```
legend("topright", legend = c("Control before", "Control after"), col = c("black", "red"), lty =
1, cex=1)
```



Interestingly, we observe the same effect in the control group which is explain by the authors for the lower real GDP after the period of the implementation of the policy and a slight increase in the benefit level for this group.

6. Estimating the causal treatment effect in a PH model

Lastly, the PH model ($\lambda_0(t)\exp(x'\beta)$) is estimated, where the baseline hazard is the object of interest.

The piecewise constant function for changes every four weeks is specified as follows:

$$\lambda_0(t)\exp\left(\sum_{l=i}^{14}\lambda_l I(4l < t < 4(l+1)) + \lambda_{15} I(t > 60)\right)$$

So, splitting the data as per usual in order to estimate the PWE PH model:

```
data %>%
  mutate(all = tr * (t39 + t52) ) ->
  data

breaks <- seq(from=3,to=59, by=4)
labels <- paste("(", c(0,breaks), ",", c(breaks,104), "]",sep="")

gux <- survSplit(Surv(dur104,uncc) ~., data=data, cut = breaks,
                 end = "time", event="death", start="start", episode="interval")

gux %>%
  mutate(exposure = time - start,
        interval=factor(interval+1, labels = labels) ) ->
  gux
```

However, I couldn't get the same results. I took out the duration, exposure and time since I noticed these had a big impact in the estimates returning totally different results.

```
library(eha)

#Define the model
model_formula = death ~ interval+.-dur-exposure-time

#Fit the model
pwe <- glm(model_formula,
           offset = log(exposure),
           data = gux,
           family = poisson)

pwe
```

```
##
## Call:  glm(formula = model_formula, family = poisson, data = gux, offset = log(exposure))
##
## Coefficients:
##      (Intercept)     interval(3,7]    interval(7,11]   interval(11,15]
##       -8.544e-01         5.976e-01         1.085e+00         1.322e+00
##  interval(15,19]   interval(19,23]   interval(23,27]   interval(27,31]
##        1.270e+00         1.369e+00         1.129e+00         1.208e+00
##  interval(31,35]   interval(35,39]   interval(39,43]   interval(43,47]
##        9.852e-01         8.382e-01         7.371e-01         5.524e-01
##  interval(47,51]   interval(51,55]   interval(55,59]  interval(59,104]
##        3.822e-01         8.643e-01         4.228e-01         3.887e-01
##           beginn           ein_zus             sfrau               age
##       -4.043e-04         3.422e-01        -4.373e-02        -1.297e-02
##            after          nwage_pj              e3_5              bdur
##       -4.871e-02        -8.817e-06         2.102e-01        -8.545e-03
##            y1988             y1989             y1990             y1991
##        1.432e-01        -3.779e-02         5.593e-02         1.353e-01
##         med_educ           hi_educ             lehre           married
##       -1.452e-01        -2.210e-01        -4.541e-02         1.482e-01
##           single          divorced            f_marr          f_single
##        1.223e-02        -9.775e-02        -7.809e-02         1.343e-01
##          f_divor                bc           pnon_10                q2
##        4.994e-02         3.734e-01         5.800e-02         5.313e-02
##               q3                q4          seasonal             manuf
##       -1.309e-01        -1.283e-01         3.283e-01        -8.844e-02
##            ten72           typePBD            typeRR       typecontrol
##       -6.358e-03         5.909e-02        -1.460e-02        -2.146e-02
##            lwage                tr               t39               t52
##        2.480e-01                NA        -3.204e-02                NA
##           t39_tr            t52_tr            after0             tr_a0
##        3.340e-02                NA                NA        -1.051e-02
##           t39_a0            t52_a0           t39tra0           t52tra0
##        7.088e-02         5.492e-01        -2.402e-02        -5.083e-01
##           after1             tr_a1            t39_a1            t52_a1
##        1.913e-01        -1.623e-01        -4.905e-02        -1.676e-01
##          t39tra1           t52tra1            after2             tr_a2
##        4.128e-02         3.076e-01        -5.713e-02         2.622e-01
##           t39_a2            t52_a2           t39tra2           t52tra2
##        2.258e-01        -7.368e-02        -3.007e-01         2.075e-01
##           after3             tr_a3            t39_a3            t52_a3
##       -2.648e-01         1.995e-01        -2.731e-02        -8.766e-02
##          t39tra3           t52tra3            after4             tr_a4
##        1.025e-01        -8.286e-02         8.065e-02        -3.566e-01
##           t39_a4            t52_a4           t39tra4           t52tra4
##       -1.221e-01         2.604e-02         3.108e-01         3.319e-03
##           after5             tr_a5            t39_a5            t52_a5
##        1.857e-01         1.159e-01        -1.638e-01        -2.815e-01
##          t39tra5           t52tra5            after6             tr_a6
##       -1.061e-01         7.135e-03        -2.091e-01         8.233e-02
##           t39_a6            t52_a6           t39tra6           t52tra6
##        5.305e-02         7.283e-02         1.154e-01         3.176e-01
```

```
##          after7           tr_a7           t39_a7           t52_a7
##       1.906e-01      -1.829e-01        5.978e-02        1.102e-03
##         t39tra7         t52tra7           after8            tr_a8
##      -1.727e-01      -1.992e-01        3.235e-01       -1.131e-01
##          t39_a8          t52_a8          t39tra8          t52tra8
##      -2.902e-02       7.403e-03        9.957e-03       -2.382e-01
##          after9           tr_a9           t39_a9           t52_a9
##      -5.049e-02       1.093e-01        6.781e-02        1.009e-01
##         t39tra9         t52tra9          after10           tr_a10
##      -9.751e-02      -3.874e-02       -1.355e-01        4.142e-03
##         t39_a10         t52_a10         t39tra10         t52tra10
##      -2.185e-02       1.430e-01        6.569e-02       -2.135e-01
##         after11          tr_a11          t39_a11          t52_a11
##       1.173e-01      -9.870e-02        2.850e-03       -1.676e-01
##        t39tra11        t52tra11          after12           tr_a12
##       3.228e-02       3.772e-01       -1.342e-01        4.072e-02
##         t39_a12         t52_a12         t39tra12         t52tra12
##       1.260e-01       2.375e-01       -2.847e-02       -2.484e-01
##         after13          tr_a13          t39_a13          t52_a13
##       6.345e-02      -1.012e-01       -1.018e-01        1.368e-03
##        t39tra13        t52tra13          after14           tr_a14
##       1.200e-01       1.550e-02        1.379e-01        7.651e-03
##         t39_a14         t52_a14         t39tra14         t52tra14
##       9.269e-03      -2.022e-01        8.882e-03        4.620e-01
##         after15          tr_a15          t39_a15          t52_a15
##      -1.088e-01       1.374e-01       -4.238e-02       -1.008e-01
##        t39tra15        t52tra15              all            start
##      -7.294e-02      -2.246e-01               NA               NA
##
## Degrees of Freedom: 1057905 Total (i.e. Null);  1057760 Residual
## Null Deviance:       1017000
## Residual Deviance: 945400    AIC: 1378000
```

```
library(stargazer)

stargazer(pwe,
          dep.var.caption="",dep.var.labels="",
          keep=1:15,
          omit.table.layout = "n", star.cutoffs = NA,
          keep.stat=c("n", "ll"),no.space=TRUE,
          header=FALSE,
          title="The PWE model", type="text"
          )
```

```
##
## The PWE model
## ============================
##
## ----------------------------
## interval(3,7]       0.598
##                    (0.009)
## interval(7,11]      1.085
##                    (0.009)
## interval(11,15]     1.322
##                    (0.009)
## interval(15,19]     1.270
##                    (0.010)
## interval(19,23]     1.369
##                    (0.011)
## interval(23,27]     1.129
##                    (0.013)
## interval(27,31]     1.208
##                    (0.014)
## interval(31,35]     0.985
##                    (0.017)
## interval(35,39]     0.838
##                    (0.019)
## interval(39,43]     0.737
##                    (0.022)
## interval(43,47]     0.552
##                    (0.025)
## interval(47,51]     0.382
##                    (0.029)
## interval(51,55]     0.864
##                    (0.025)
## interval(55,59]     0.423
##                    (0.032)
## interval(59,104]    0.389
##                    (0.015)
## ----------------------------
## Observations      1,057,906
## Log Likelihood   -689,066.600
## ============================
```

So, by doing this, I got closer results as the ones you had.

Conclusion

Due to unemployment insurance reforms across many countries, this study delves into the critical issue of how financial incentives within the system impact the duration of unemployment. It analyses the combined effects of changes in different unemployment insurance parameters, an area where previous research often lacked detail, making it difficult to isolate the relative impact of specific policy changes.

From a policy perspective, the study suggests that potential benefit duration (PBD) emerges as a more effective tool than the level of benefits (RR) in influencing job search behavior and unemployment durations. This implies that policymakers might achieve better results by focusing on the overall duration of unemployment spells rather than just the duration of benefits. Additionally, the research emphasizes the need to consider behavioral effects and the relationship between different incentives and institutional settings when designing unemployment insurance policies.

The replication of this study using survival analysis revealed that the new policies did not significantly reduce unemployment spells; in fact, they might even have extended them due to the potential disincentive to job search created by longer benefit durations. This reinforces the argument for focusing on measures that directly address unemployment duration itself rather than solely on the length of benefit periods.

In conclusion, this study offers valuable insights into the intricate relationship between unemployment benefit parameters and job search behavior. By highlighting the importance of considering both the effects of individual parameters and their interactions, the research equips policymakers with the knowledge needed to design unemployment insurance systems that better balance social support with the goal of facilitating re-employment and reducing unemployment durations.