Assessment
(The Duration of Unemployment - Replicating Lalive et al. (2006, ReStud))

1 Rules and Logistic Details

This **assessed** homework (i.e. your research report) will contribute **50%** to your final grade for this course. The deadline for submission is **Wednesday 12.12.2018 at 12:00** (but you can submit it, of course, before). The submission consists of a printout of a **notebook** (such as a Rstudio or Jupyter notebook), in which you combine discussions of questions as well as computations.

The overall objective of this exercise is to 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.

(all necessary materials, such as the paper, the data set, as well as a complete set of benchmark coefficients, are posted on Ametice)

2 Research objectives

Q1 **What are the contributions of this study ?**

Write a brief summary (no more than 1.5 pages) that discusses the contributions of this study. You might wish to consider the following questions: What are the objectives, and how do they relate to the established literature? What is novel about the empirical strategy? What are the results and how convincing are these?

3 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: 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.

4 Data Preparation

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.

The duration of interested is right censored at 104 weeks, $t_u^{104} = \min\{t_u, 104\}$, which we have to generate. The exit indicator is `uncc`, the groups are defined by the variable `type` which is coded as a factor (levels can be examined using `levels(udat$type)`). The binary variable `after` indicates the period after the policy change.

```
> rm(list=objects())
> library(foreign)
> library(survival)
> udat <- read.dta("fi.dta")  ## size: 150 MB !
> dim(udat)  ## 225,821 by 166
[1] 225821    166

> udat <- udat[,1:134] ## get rid of some superfluous variables

> table(udat$type)
PBD and RR        PBD         RR    control
     21174      99404      32470      72773

> ## Computation of average spells when durations are truncated at 104 weeks
> udat$dur104 <- udat$dur
> udat$dur104[(udat$dur104 > 104)] = 104
```

The RDD design is as follows: Before 1 August, 1989: PBD was 20 weeks, RR about 41%. After 1 August, 1989: RR increased to about 47%. PBD became dependent not only on previous contributions but also on *age* at the beginning of the unemployment spell. Benefit duration for the age group 40-49 was increased to 39 weeks (if the unemployed has been within the last 10 years prior to the current spell). For the age group 50 and older, PBD was increased to 52 weeks.

The variables and interaction terms contained in the dataset code this design, see Table 1.

Table 1: Variable definitions

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
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

Q2 (difference-in-differences) **Attempt to replicate Table 4 of the paper**. As it will turn out that there are small discrepancies, Table 2 reports the results.

Recall that the difference-in-differences estimator is:

$$\tilde{\Delta}_{DD} = \left(\bar{Y}_A^T - \bar{Y}_B^T\right) - \left(\bar{Y}_A^C - \bar{Y}_B^C\right)$$

where $\bar{Y}_B^T$ and $\bar{Y}_A^T$ are the average duration of unemployment for the treated group before and after the date of the reform. $\bar{Y}_B^C$ and $\bar{Y}_A^C$ are the average duration of unemployment for the control group before and after the reform.

Q3 (Survival Functions) **Seek to reporduce Figure 3** in Lalive et al. (2006).

Lalive et al. (2006) also report in Figure 4 hazard estimates based on the KM estimates. It is not quite clear how these have been generated, since these figures

Table 2: Difference-in-Difference estimates, Table 4 in paper.

	Before August 1989	After August 1989	Change (after-before)	Diff-in-diff (compared to control)
ePBD group	16.25 (0.08)	18.67 (0.09)	2.42 (0.12)	1.13 (0.18)
N	48,294	51,110		
eRR group	17.79 (0.12)	20.03 (0.16)	2.24 (0.20)	0.96 (0.24)
N	17,160	15,310		
ePBD-RR group	19.01 (0.17)	23.55 (0.24)	4.53 (0.20)	3.25 (0.24)
N	11,992	9,182		
Control group	15.24 (0.08)	16.52 (0.09)	1.29 (0.13)	
N	33,815	38,958		

are very smooth. One way of proceeding it to estimate the hazard as

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

If we use the discrete approximation to f based on $\hat{S}(t)$, this will be very erratic. One way to increase smoothness is to pick a smaller number of points on $\hat{S}(t)$, e.g. one for each week. It turns out that the result is still quite unsmooth. To increase smoothness further, one can pass a smoother of the curve, such as a local polynomial regression smoother. One such smoother, the `locpoly` function, is found in the `KernSmooth` library.

Q4 (KM estimates of the unemployment exit hazard) **Seek to reporduce Figure 4** in Lalive et al. (2006). Using `locpoly`, my smoothed version is reproduced in Figure 1 as an illustration.

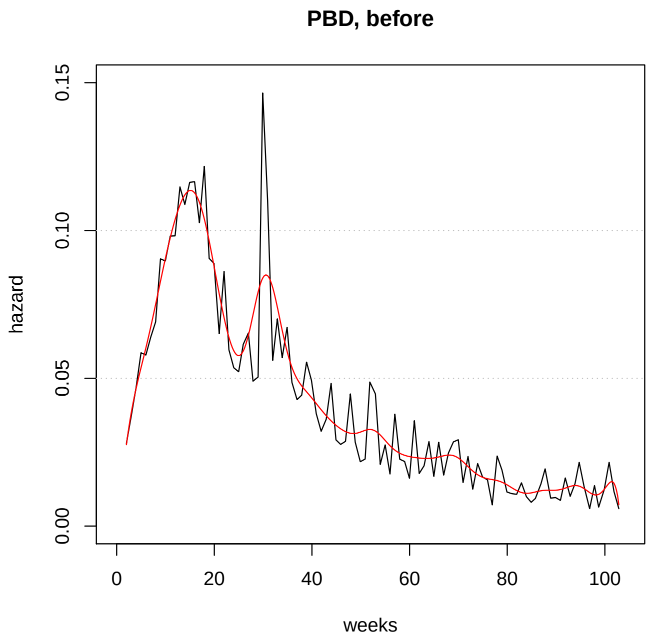
The authors estimate a PH model, $\lambda_0(t)\exp(x'\beta)$, where the baseline hazard is the key object of interest (since it refers to the exit rate for a homogeneous group of workers). Specifically, it is specified as a piecewsie constant function, where shift can occur in every four-week interval.

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

The trick is in specifying the coefficients λ so as to allow policy changes to have an effect

$$\begin{aligned}
\lambda_l = \beta_{0l} &+ \beta_{1l}\mathbf{eP39} + \beta_{2l}\mathbf{eP52} + \beta_{3l}\mathbf{eRR} \\
&+ \beta_{4l}(\mathbf{eP39} + \mathbf{eP52}) * \mathbf{eRR} + \beta_{5l}A89 \\
&+ \delta_{1l}\mathbf{eP39} * A89 + \delta_{2l}\mathbf{eP52} * A89 + \delta_{3l}\mathbf{eRR} * A89 \\
&+ \delta_{4l}\mathbf{eP39} * \mathbf{eRR} * A89 + \delta_{5l}\mathbf{eP52} * \mathbf{eRR} * A89
\end{aligned}$$

Figure 1: Hazard Functions



The δs measure the change in duration dependence of the hazard rate due to changes in financial incentive 6 parameters because the policy change entails five interventions. E.g. the interaction term eP39*A89(tc) indicates that an individual satisfying all eligibility criteria for the extension to 39 weeks has entered the period when this policy change has been enacted.

If your computer has insufficient RAM, take a random sample before preparing the PH model estimation.

```
i = as.integer(length(udat$dur)*runif(1000)) ## NB: sample(.) only works on vectors
udat <- udat[i,]
```

Then split the data as per usual in order to estimate the PWE PH model

```
udat$all <- udat$tr * (udat$t39 + udat$t52)

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

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

library(dplyr) ## for mutate function
gux <- mutate(gux,exposure = time - start,
              interval=factor(interval+1, labels = labels) )
```

Q5 **Estimate the causal treatment effect in a PH model.** The coefficient estimates are reported by Lalive et al. in the separately posted document entitled `lalive_coeff_estimates_PH_appendixtable.pdf`.