Assessing Causal Effects in a longitudinal observational study with “truncated” outcomes due to unemployment and nonignorable missing data

Michela Bia [1]

Alessandra Mattei [2]

Andrea Mercatanti [3]

ABSTRACT

In this paper we analyze the short- and long-run effect of foreign language training programs on employment and wages measured over time, using administrative data on labour force in Luxembourg (IGSS-ADEM dataset). We develop a novel framework to simultaneously handle truncated wages due to unemployment, with incomplete observations not ignorable over time. In our study we find that language training programs increased re-employment probabilities, with no effect on the wages. This might be an incentive for the Employment Agency to better design future policies implemented in the context of language trainings. We then focus the analysis on the group of defiant-employees and find that defiers at 18 months switch to the always-employees stratum at 36 months with a proportion of almost 50% (the highest transition probability between the two periods). This evidence is in line with the economic theory: defiant-employees are subjects who accept any job, when not trained, but prefer to wait for a

[1]Evaluation Unit, Labour Market Department, LISER, Luxembourg. Email: michela.bia@liser.lu.

Michela Bia acknowledges financial support from the European Social Fund Project: “Evaluation of Active Labor Market Policies in Luxembourg” – EvaLab4Lux, cofunded by the Ministry of Labour, Employment and the Social and Solidarity Economy of Luxembourg and Liser.

[2]Department of Statistics, Computer Science, Applications, University of Florence, Italy. Email: mattei@disia.unifi.it

[3]Evaluation Unit, Labour Market Department, LISER, Luxembourg and Bank of Italy, Rome. Email: andrea.mercatanti@liser.lu

job with higher wage, when exposed to the program, because they feel better equipped.

KEY WORDS: principal stratification, propensity score, unconfoundedness, training programs, censored outcomes, longitudinal data, reservation wage, defiers.

1 Introduction

In this paper we analyze the short- and long-run effect of a language training program on subsequent labour market outcomes (employment and wages) measured over time using data from an observational study conducted in Luxembourg. Specifically, the dataset we use is obtained merging the rich administrative database on workers' trajectories provided by the Luxembourg's global security database on labour force (IGSS), with the information collected by the employment Agency (ADEM). These data were previously analyzed by Bia et al. (2017), who focused on causal effects of the training program on employment and hourly wage at 18 months after registering at the Employment Agency. Here we are also interested in assessing how causal effects on employment and hourly wage evolve over time, so we use longitudinal information on employment and hourly wage at 18 months and 36 months after entering the unemployment status.

The study suffers from a number of complications, that make the evaluation analysis particularly challenging. First, it is an observational study, so some assumptions on the assignment mechanism is required (e.g., Imbens and Rubin (2015)). Second, hourly wages are "truncated by death", because they are neither observed nor defined for subjects who are unemployed in a given point in time (Rubin (2000), Rubin (2006); Zhang and Rubin (2003); Zhang et al. (2008), Frumento et al. (2012)). Third, employment and hourly wages are missing for some participants in the study at the various points in time and the missingness may be nonignorable, because it is related to the not fully measured outcomes (Rubin (1976), Little and Rubin (2002); Mattei et al. (2014); Mealli and Rubin (2015)). In fact missing values in the longitudinal data occur when individuals either become inactive or leave the country (for example because they find a job abroad). In our data, about 15% and 20% of the subjects have missing values at 18 months and at 36 months, respectively.

In our study we focus on describing and addressing these complications under the potential outcome approach to causal inference using the framework of principal stratification (PS) (Frangakis and Rubin, 2002) using a model-based Bayesian mode of inference.

Following Bia et al. (2017), we design the observational study under the assumption of unconfoundedness, which rules out the presence of unmeasured confounders conditional on the observed covariates. Noteworthy, the unconfoundedness assumption appears to be reasonable in our study: Bia et al. (2017) also conducted a sensitivity analysis to account for unobserved confounding and found that the estimated results were robust to departures from uncounfoundedness assumptions.

We address the problem of wages truncated by unemployemnt using the framework of PS, which is, nowadays, widely adopted in the evaluation of public policies (e.g., Zhang et al. (2008), Frumento et al. (2012), Mattei et al. (2013)) in that it allows to properly adjust for post-treatment variables, that may be affected by the treatment. In our setting, at each point in time, principal stratification classifies subjects into four (latent) groups with respect to the joint potential values of the employment status under each treatment condition: always-employees (who would be employed regardless of their treatment assignment); never-employees (who would be unemployed regardless of their treatment assignment); compliant-employees (who would be employed only if assigned to the language program); and defiant-employees (who would be employed only if assigned to the control group). This stratification of units makes it clear that causal effects on wages are defined only for always-employees.

Finally, we deal with the presence of nonignorable nonresponse on both the intermediate outcome (employment) and the primary outcome (wages), using an ad hoc model for the missing data process, where we allow the missingness to depend on the partially observed strata defined by the employment status.

In our case study, the types of subjects defined by principal stratification with respect to the employment status deserve some discussion. Although causal effects on wages are only defined for always-employees, in our study principal strata of particularly interest also comprise defiant-employees. Defiant-employees might be subjects who would accept any job, when not trained, but would prefer to wait for a better job (with higher wage) when exposed to the program. Even more than compliant-employees, defiant-employees could reasonably think that the intervention improved their job skills, leading to a substantial increase in their reservation wage, the lowest wage at which an individual is willing to accept a job. Specifically, defiant-employees formerly unemployed after receiving the training, because of a rise in their reservation wage, might be induced to accept a work later (for example in 2 or 3 years), when their potential wage surpasses (at a certain point in time) their reservation wage. Such a behavior would give evidence about the existence of a positive effect on sub-groups of people different to the ones usually considered in this type of analyses, providing policy makers with additional causal statements useful to optimally design future active labor market policies (ALMP).

The concept of reservation wage has been central in many theoretical works on models of job search, labour supply and labour market participation (e.g., Mortensen (1986); Jones (1988)). In this context, defiant-employees' behavior has never been studied in the existing economic literature, mainly focused on the effect of unemployment benefits and nonemployment duration on reservation wages (a seminal work is from Lancaster and Chesher (1983) and recent contributions are from Schmieder et al. (2016), Krueger and Mueller (2016), and Brown and Taylor (2013)).

In a recent study on the effect of ALMP, Sørensen (2017) shows that ALMPs deliver heterogeneous effects on earnings and argues that positive effects might be driven by either a faster return to employment together with a lowering of reservation wages or a more moderate return to employment together with an increase in reservation wages. Using data from a randomized experiment conducted in two Danish counties during 2005-2006, he shows that treated individuals are more likely to gain formal human capital accumulation, and hence raising their reservation wages. Conversely, using the same data, negative effects on the reservation wage (and positive effects on search effort) are found when studying the 'threat' effect of an active labor market policy regime (Rhosholm and Svarer (2008); see also the work by Van den Berg et al. (2009) on the German system of active labor market policies).

The program evaluation literature focusing on assessing causal effects on wage in the presence of unemployment has mainly focused on the principal average causal effect for always-employees (e.g., Zhang et al. (2009)), neglecting the other principal strata. An exception is the work by Frumento et al. (2012) who proposed to characterize each latent subgroup (not only the principal stratum of always-employees) in terms of its background characteristics.

Our contribution to the existing literature si twofold: first, we advance the methodology under the principal stratification model by introducing a novel framework to simultaneously handle truncated labor market outcomes due to unemployment, with incomplete observations not ignorable over time. Second, we fill the gap in the empirical studies on ALMPs by providing informative evidence on defiant-employees' behavior, which is possible to follow over time by exploiting the longitudinal structure of the data.

Defiers formerly unemployed after receving the training, because of a rise in their reservation wage, might be induced to work later (for example in 2 or 3 years), when their potential wage surpasses (at a certain point in time) their reservation wage. Such a behavior would give evidence about the existence of a positive effect on sub-groups of people different to the ones usually considered in this type of analyses, providing policy makers with additional causal statements useful to optimally design future ALM policies.

The remainder of the paper is organized as follows. Section 2 and 3 introduce the basic theoretical framework and the causal model, respectively. In section 4 we describe the data used in our study and section 5 presents some preliminary results. Finally, section 6 concludes.

2 Basic theoretical framework and notation

We adopt the Potential Outcome Approach to causal inference throughout the paper. We have a sample of $i = 1, ..., N$ unemployed individuals for which we observe k pre-treatment variables $\mathbf{X}_i$, and the binary treatment Z_i which is equal to C if i does not participate in a language course at ADEM (Control unit), equal to T otherwise (Treated unit). We define also the following quantities, potentially observed at two times under each value of the treatment Z, time $t = 18, 36$ months after entering unemployment:

- a binary post-treatment variable $S(Z)_{i,t}$ which indicates the i's employment status at t, equal to 0 if i is still unemployed at t, to 1 if employed;
- the continuous outcome $Y(Z)_{i,t}$ which is the log of hourly wage for i at t if $S(Z)_{i,t} = 1$;
- the binary response indicator $R(Z)_{i,t}$ which indicates whether, at t, the post-treatments quantities $S(Z)_{i,t}$ and $Y(Z)_{i,t}$ are missing $R(Z)_{i,t} = 0$, or not $R(Z)_{i,t} = 1$. When $R(Z)_{i,t} = 0$ then $S(Z)_{i,t}$ and $Y(Z)_{i,t}$ are both missing and we pose $S(Z)_{i,t} = (Z)_{i,t} =?$.

We arrange the aforementioned variables in the following vectors and matrices:

- $\mathbf{X}$: $N \times k$ matrix of pre-treatment variables with $\mathbf{X}_i$ being the i-row vector.
- $\mathbf{Z}$: $N \times 1$ vector of the individual treatments $Z_i = C, T$.
- $\underline{\mathbf{S}}_t$: $N \times 2$ matrix with $\underline{S_{i}}_t \equiv (S(C)_{i,t}, S(T)_{i,t})$ being the i-row vector
- $\underline{\mathbf{Y}}_t$: $N \times 2$ matrix with $\underline{Y_{i}}_t \equiv (Y(C)_{i,t}, Y(T)_{i,t})$ being the i-row vector
- $\underline{\mathbf{R}}_t$: $N \times 2$ matrix with $\underline{R_{i}}_t \equiv (R(C)_{i,t}, R(T)_{i,t})$ the i-row vector

For each of the aforementioned couples of potential quantities, $S(Z)_{i,t}$, $Y(Z)_{i,t}$, $R(Z)_{i,t}$, with $Z = T, C$, we observe only the quantity corresponding to the actual value of Z. In order to account for the selection into the treatment, because individuals are not randomly assigned to the treatment, we invoke the following ignorability assumption (Rosenbaum and Rubin (1983); Bia et al. (2017)):

Assumption 1 *(Strong Ignorability)*

– *Unconfondedness:* $Z_i \perp \underline{\mathbf{R}}_{18}, \underline{\mathbf{S}}_{18}, \underline{\mathbf{Y}}_{18}, \underline{\mathbf{R}}_{36}, \underline{\mathbf{S}}_{36}, \underline{\mathbf{Y}}_{36} \mid \mathbf{X}_i$

– *Overlap:* $0 < Pr(Z_i = 1 \mid \mathbf{X}_i)$

We deal with the selection into the outcome, which arises because wages are defined only for individuals who re-employ, we will use the principal stratification framework (Zhang et al. (2009); Frumento et al. (2012)). $\underline{S_{i_t}}$, which is the couples of values of the post-treatment variables under each value of the treatment Z, define the principal stratum, $G_{i,t}$. Because of the binary nature of the the treatment Z and the intermediate outcome S, we have four principal stratum at t:

- Always-employed, which are subjects who would be employed regardless of treatment assignment: $G_{i,t} = EE$ when $\underline{S_{i_t}} \equiv (S(C)_{i,t}, S(T)_{i,t}) = (1,1)$
- Never-employed, which are subjects who would not be employed regardless of treatment assignment: $G_{i,t} = NN$ when $\underline{S_{i_t}} \equiv G_{i,t} \equiv (S(C)_{i,t}, S(T)_{i,t}) = (0,0)$
- Compliers, which are subjects who would be employed under treatment, but not employed under control: $G_{i,t} = NE$ when $\underline{S_{i_t}} \equiv G_{i,t} \equiv (S(C)_{i,t}, S(T)_{i,t}) = (0,1)$
- Defiers, which are subjects who would not be employed under treatment but employed under control: $G_{i,t} = EN$ when $\underline{S_{i_t}} \equiv G_{i,t} \equiv (S(C)_{i,t}, S(T)_{i,t}) = (1,0)$

3 The causal model

We adopt the phenomenological Bayesian approach to causal inference as introduced by Rubin (1978) and extended to the framework of a randomized experiment with non-compliance by Imbens and Rubin (1997). The latter is a special case of principal strata model where the intermediate variable is the actual taking of a randomly assigned treatment; it differs from our framework where randomization is supposed to hold only conditionally to a set of pre-treatment variables and where there are not non-compliance issues. However, Imbens and Rubin (1997) is a suitable starting point for a more complex model where the outcomes are repeatedly observed over time and where attrition can occur.

We retain plausible the probability of missingness in the repeated outcomes may depend on the values of the outcomes themselves, therefore we develop an ad hoc model for the missing data process (Little and Rubin (2002); Schafer and Graham (2002)). Formally we make the following assumption:

Assumption 2 *For each $Z = C, T$*

$$R(T)_{i,18} \perp R(C)_{i,18}, Y_(C)i, 18, Y_(T)i, 18, \mid \mathbf{X}_i, G_{i,18}$$

$$R(T)_{i,36} \perp R(C)_{i,18}, Y_(C)i, 18, Y_(T)i, 18, Y_(C)i, 36, Y_(T)i, 36 \mid \mathbf{X}_i, R(Z)_{i,18}, G_{i,36}$$

Note that in our study, if $R(Z)_{i,18} = 0$, then $R(Z)_{i,36} = 0$

Under Assumptions 1 and 2 we propose a causal model articulated taking into account two sub-models: one for the complete data (namely the union of the observed and the missing data) and another one for the response indicator given the complete data. This way we will be able to obtain the model for the complete data, from which the model for the osbervable data can be derived by integrating out the missing quantities. The model for the complete data can be formalized in term of potential quantities (Rubin (1978); Imbens and Rubin (1997)) as follow:

$$f(\mathbf{X}, \mathbf{Z}, \underline{\mathbf{S}}_{18}, \underline{\mathbf{Y}}_{18}, \underline{\mathbf{R}}_{18}, \underline{\mathbf{S}}_{36}, \underline{\mathbf{Y}}_{36}, \underline{\mathbf{R}}_{36}) \propto f(\mathbf{Z}) \cdot f(\underline{\mathbf{S}}_{18}, \underline{\mathbf{Y}}_{18}, \underline{\mathbf{R}}_{18}, \underline{\mathbf{S}}_{36}, \underline{\mathbf{Y}}_{36}, \underline{\mathbf{R}}_{36} | \mathbf{X}).$$

By assuming exchangeability and appealing to the De Finetti's theorem, we can assume the units to be independent and identically distributed given the parameter vector $\boldsymbol{\pi}$ with prior distribution $P(\boldsymbol{\pi})$, so that:

$$f(\underline{\mathbf{S}}_{18}, \underline{\mathbf{Y}}_{18}, \underline{\mathbf{R}}_{18}, \underline{\mathbf{S}}_{36}, \underline{\mathbf{Y}}_{36}, \underline{\mathbf{R}}_{36} | \mathbf{X}) = \int \prod_i f(\underline{S_i}_{18}, \underline{Y_i}_{18}, \underline{R_i}_{18}, \underline{S_i}_{36}, \underline{Y_i}_{36}, \underline{R_i}_{36} | \mathbf{X}_i, \boldsymbol{\pi}) P(\boldsymbol{\pi}) d\boldsymbol{\pi}$$

Now, let's indicate with $(\underline{S_{i,obs,t}}, \underline{Y_{i,obs,t}}, \underline{R_{i,obs,t}}, \underline{\mathbf{S}_{obs,t}}, \underline{\mathbf{Y}_{obs,t}}, \underline{\mathbf{R}_{obs,t}})$ the observed potential quantities, namely the manifestation of the potential quantities under the actual value of the treatment, and with $(\underline{S_{i,mis,t}}, \underline{Y_{i,mis,t}}, \underline{R_{i,mis,t}}, \underline{\mathbf{S}_{mis,t}}, \underline{\mathbf{Y}_{mis,t}}, \underline{\mathbf{R}_{mis,t}})$ the unobserved potential quantities, namely the manifestation of the potential quantities under the value of the treatment not assigned. We adopt the aforementioned underlined notation to distinguish the missingness due to the concept of potential quantities from the missingness due to attrition. For example, in the following, $\underline{S_{i,mis,t}}$ is the missing potential outcomes for employment status for i, while $S_{i,mis,t}$ is the actual, but missing (due to nonresponse), employment status for i (for example because i is expatriated).

The posterior distribution of $\boldsymbol{\pi}$ can be formalized as:

$$\begin{aligned} &P(\boldsymbol{\pi} | \mathbf{X}, \mathbf{Z}, \mathbf{S}_{obs,18}, \mathbf{Y}_{obs,18}, \mathbf{R}_{obs,18}, \mathbf{S}_{obs,36}, \mathbf{Y}_{obs,36}, \mathbf{R}_{obs,36}) \propto \\ &\propto P(\boldsymbol{\pi}) \sum_{\underline{\mathbf{S}_{mis,18}}} \sum_{\underline{\mathbf{S}_{mis,36}}} \int \cdots \int \sum_{\underline{\mathbf{R}_{mis,18}}} \sum_{\underline{\mathbf{R}_{mis,36}}} f(\underline{\mathbf{S}}_{18}, \underline{\mathbf{Y}}_{18}, \underline{\mathbf{R}}_{18}, \underline{\mathbf{S}}_{36}, \underline{\mathbf{Y}}_{36}, \underline{\mathbf{R}}_{36} | \mathbf{X}, \boldsymbol{\pi}) d\underline{\mathbf{Y}_{mis,18}} d\underline{\mathbf{Y}_{mis,36}} = \\ &= P(\boldsymbol{\pi}) \prod_i \sum_{\underline{S_{i,mis,18}}} \sum_{\underline{S_{i,mis,36}}} \int \cdots \int \sum_{\underline{R_{i,mis,18}}} \sum_{\underline{R_{i,mis,36}}} f(\underline{S_i}_{18}, \underline{Y_i}_{18}, \underline{R_i}_{18}, \underline{S_i}_{36}, \underline{Y_i}_{36}, \underline{R_i}_{36} | \mathbf{X}_i, \boldsymbol{\pi}) d\underline{Y_{i,mis,18}} d\underline{Y_{i,mis,36}} \end{aligned} \quad (1)$$

where the summation and integration operators act on the unobserved potential quantities. Let's pose $\boldsymbol{\pi} = (\boldsymbol{\omega}_{18\cap 36}, \boldsymbol{\eta}_{gC,18\cap 36}, \boldsymbol{\eta}_{gT,18\cap 36}, \boldsymbol{\eta}_{gCT,18\cap 36})$ where $\boldsymbol{\omega}_{18\cap 36}$ is the parameter vector for the principal strata model, $\boldsymbol{\eta}_{gC,18\cap 36}$ the parameter vector of the joint model for potential outcomes for the response indicator and wage under control, $\boldsymbol{\eta}_{gT,18\cap 36}$ the parameter vector of the joint model for potential outcomes for the response indicator and wage under treatment, and $\boldsymbol{\eta}_{gCT,18\cap 36}$ the association parameter vector. Let's introduce also the indicator $\delta(\mathbf{g}, \underline{S_{i}}_{18}, \underline{S_{i}}_{36})$ which is equal to 1 if $(\underline{S}_{i18}, \underline{S}_{i36})$ implies $\mathbf{g} \equiv (G_{i18}, G_{i36})$ and equal to 0 otherwise.

We can write

$$
\begin{gathered}
f(\underline{S_{i}}_{18}, \underline{Y_{i}}_{18}, \underline{R_{i}}_{18}, \underline{S_{i}}_{36}, \underline{Y_{i}}_{36}, \underline{R_{i}}_{36}|\mathbf{X}_i, \pi) = \\
= \sum_{\mathbf{g}} \delta(\mathbf{g}, \underline{S_{i}}_{18}, \underline{S_{i}}_{36}) P(\mathbf{g}|\mathbf{X}_i, \boldsymbol{\omega}_{18\cap 36}) \times f_{\mathbf{g}C}(Y(C)_{i,18}, Y(C)_{i,36}, R(C)_{i,18}, R(C)_{i,36}|\mathbf{X}_i, \mathbf{g}, C, \boldsymbol{\eta}_{\mathbf{g}C,18\cap 36}) \times \\
\times f_{\mathbf{g}T}(Y(T)_{i,18}, Y(T)_{i,36}, R(T)_{i,18}, R(T)_{i,36}|\mathbf{X}_i, \mathbf{g}, T, \boldsymbol{\eta}_{\mathbf{g}T,18\cap 36}) \times h_{\mathbf{g}}(\underline{Y_{i}}_{18}, \underline{R_{i}}_{18}, \underline{Y_{i}}_{36}, \underline{R_{i}}_{36}|\mathbf{X}_i, \mathbf{g}, \boldsymbol{\eta}_{\mathbf{g}CT,18\cap 36})
\end{gathered}
$$

where $h_{\mathbf{g}}(\underline{Y_{i}}_{18}, \underline{R_{i}}_{18}, \underline{Y_{i}}_{36}, \underline{R_{i}}_{36}|\mathbf{X}_i, \mathbf{g}, \boldsymbol{\eta}_{gCT,18\cap 36})$ is defined such that

$f_{\mathbf{g}C} \times f_{\mathbf{g}T} \times h_{\mathbf{g}} = f(\underline{Y_{i}}_{18}, \underline{R_{i}}_{18}, \underline{Y_{i}}_{36}, \underline{R_{i}}_{36}|\mathbf{X}_i, \mathbf{g}, \boldsymbol{\pi})$.

The sums and integrations in (1) lead to a mixture structure for the posterior where the number of mixtures is equal to the number of the terns $(Z_{i,obs}, S_{i,obs,18}, S_{i,obs,36})$. For example, if $(Z_{i,obs} = C, S_{i,obs,18} = 0, S_{i,obs,36} = 1)$ the sum over $\underline{S_{i,mis,t}} = S(T)_{i,t}$, $t = 18, 36$, and the indicator $\delta(\mathbf{g}, \underline{S_{i}}_{18}, \underline{S_{i}}_{36})$ eliminate always-employees and defiant-employees at $t = 18$, other than compliant-employees and never-employees at $t = 36$. Further integrations over $\underline{Y_{i,mis,18}} = Y(T)_{i,18}$ and $\underline{Y_{i,mis,36}} = Y(T)_{i,36}$ as well as the sums on $\underline{R_{i,mis,18}} = R(T)_{i,18}$ e $\underline{R_{i,mis,36}} = R(T)_{i,36}$ eliminate the remaining factors which involve $f_{\mathbf{g}T}(\cdot)$ e $h_{\mathbf{g}}(\cdot)$, since:

$$
\begin{gathered}
h_{\mathbf{g}}(\underline{Y_{i}}_{18}, \underline{R_{i}}_{18}, \underline{Y_{i}}_{36}, \underline{R_{i}}_{36}|\mathbf{X}_i, \mathbf{g}, \boldsymbol{\eta}_{gCT,18\cap 36}) = \\
= f_{\mathbf{g}C}(Y(C)_{i,18} \cdots)^{-1} \times f_{\mathbf{g}T}(Y(T)_{i,18} \cdots)^{-1} \times f(\underline{Y_{i}}_{18}, \underline{R_{i}}_{18}, \underline{Y_{i}}_{36}, \underline{R_{i}}_{36}|\mathbf{X}_i, \mathbf{g}, \boldsymbol{\pi})
\end{gathered}
$$

and

$$\int\int \sum_{\underline{R(T)_{i,18}}} \sum_{\underline{R(T)_{i,36}}} f_{\mathbf{g}C}(Y(C)_{i,18}, \cdots) \times f_{\mathbf{g}T}(Y(T)_{i,18}, \cdots) \times h_{\mathbf{g}}(\underline{Y_i}_{18}, \cdots) d\underline{Y(T)_{i,18}} d\underline{Y(T)_{i,36}} =$$

$$= f_{\mathbf{g}C}(Y(C)_{i,18} \cdots) \int\int \sum\sum f_{\mathbf{g}T}(Y(T)_{i,18} \cdots) \times h_{\mathbf{g}}(\underline{Y_i}_{18}, \cdots) =$$

$$= f_{\mathbf{g}C}(Y(C)_{i,18} \cdots) \int\int \sum\sum f_{\mathbf{g}C}(Y(T)_{i,18} \cdots |\boldsymbol{\eta}_{gC,18\cap 36})^{-1} \times f(\underline{Y_i}_{18}, \cdots) =$$

$$= f_{\mathbf{g}C}(Y(C)_{i,18} \cdots) \int\int \sum_{\underline{R(T)_{i,18}}} \sum_{\underline{R(T)_{i,36}}} f(Y(T)_{i,18}, Y(T)_{i,36}, R(T)_{i,18}, R(T)_{i,36}|$$

$$|Y(C)_{i,18}, Y(C)_{i,36}, R(C)_{i,18}, R(C)_{i,36}, \mathbf{X}_i, \mathbf{g}) d\underline{Y(T)_{i,18}} d\underline{Y(T)_{i,36}} = f_{\mathbf{g}C}(Y(C)_{i,18} \cdots)$$

Therefore for $i \in (Z_{i,obs} = C, S_{i,obs,18} = 0, S_{i,obs,36} = 1)$ we obtain the mixture

$$P(NE, EE) f_{(NE,EE)C}(Y(C)_{i,18} \cdots) + P(NE, EN) f_{(NE,EN)C}(Y(C)_{i,18} \cdots) +$$

$$+ P(NN, EE) f_{(NN,EE)C}(Y(C)_{i,18} \cdots) + P(NN, EN) f_{(NN,EN)C}(Y(C)_{i,18} \cdots).$$

We have shown the likelihood for the complete data in (1) has a mixture structure. Further sums and integrations over the missing outcomes $S_{i,mis,t}$ and $Y_{i,mis,t}$ lead to the posterior distribution for the observed data:

$$\begin{aligned} P(\boldsymbol{\pi}) \prod_i \Big[& I(Y_{i,18} \neq ?, Y_{i,36} \neq ?) f(\cdot) + \\ & + I(Y_{i,18} = ?, Y_{i,36} \neq ?) \sum_{S_{i,18}} \int f(\cdot) dY_{i,18} + \\ & + I(Y_{i,18} \neq ?, Y_{i,36} = ?) \sum_{S_{i,36}} \int f(\cdot) dY_{i,36} + \\ & + I(Y_{i,36} = ?, Y_{i,36} = ?) \sum_{S_{i,18}} \sum_{S_{i,36}} \int\int f(\cdot) dY_{i,18} dY_{i,36} \Big] \end{aligned} \tag{2}$$

where

$$f(\cdot) = \sum_{\underline{S_{i,mis,18}}} \sum_{\underline{S_{i,mis,36}}} \int \cdots \int \sum_{\underline{R_{i,mis,18}}} \sum_{\underline{R_{i,mis,36}}} f(\underline{S_{i}}_{18}, \underline{Y_{i}}_{18}, \underline{R_{i}}_{18}, \underline{S_{i}}_{36}, \underline{Y_{i}}_{36}, \underline{R_{i}}_{36} | \mathbf{X}_i, \boldsymbol{\pi})$$

$$d\underline{Y_{i,mis,18}} d\underline{Y_{i,mis,36}}$$

Taking into account the constraint $R(z)_{i,36} = 0$ if $R(z)_{i,18} = 0$, we define the following models:

- for the strata:

  $P(G_{i,18\cap 36}|\mathbf{X}_i, \boldsymbol{\omega}_{18\cap 36}) = P(G_{i,18}|\mathbf{X}_i, \boldsymbol{\omega}_{18}) P(G_{i,36}|\mathbf{X}_i, G_{i,18}, \boldsymbol{\omega}_{36})$,

- for the outcomes and the response indicators under control

  $f_{\mathbf{g}C}(Y(C)_{i,18}, Y(C)_{i,36}, R(C)_{i,18}, R(C)_{i,36}|\mathbf{X}_i, \mathbf{g}, C, \boldsymbol{\eta}_{\mathbf{g}C,18\cap 36}) = f(Y(C)_{i,18}|\mathbf{X}_i, G_{i,18}, C, \boldsymbol{\eta}^Y_{gC,18})\cdot$
  $f(R(C)_{i,18}|\mathbf{X}_i, G_{i,18}, C, \boldsymbol{\eta}^R_{gC,18}) \cdot f(Y(C)_{i,36}|\mathbf{X}_i, G_{i,36}, C, \boldsymbol{\eta}^Y_{gC,36}) \cdot$
  $f(R(C)_{i,36}|\mathbf{X}_i, R(C)_{i,18}, G_{i,36}, C, \boldsymbol{\eta}^R_{gC,36})$,

  analogous formulation holds for the outcomes and the response indicators under treatment.

In particular, we pose:

- the strata at $t = 36$ depending on the strata at the previous time $t = 18$:

  $P(G_{i,18} = g|\mathbf{X}_i, \boldsymbol{\omega}_{18})$: multiLogit$(\alpha^{G18}_{g,18} + \mathbf{X}_i^T \boldsymbol{\beta}^{G18}_{g,18})$,

  $P(G_{i,36} = g|\mathbf{X}_i, G_{i,18}, \boldsymbol{\omega}_{36})$: multiLogit$(\alpha^{G36}_{g,36} + \mathbf{X}_i^T \boldsymbol{\beta}^{G36}_{g,36} + \gamma^G_{g,NE} NE_{i,18} + \gamma^G_{g,EE} EE_{i,18} +$
  $\gamma^G_{g,NN} NN_{i,18})$,

  where $g \in \{NE, NN, EE, EN\}$, and defiers are the baseline $\alpha^{G18}_{EN,18} = \boldsymbol{\beta}^{G18}_{EN,18} = 0$

- the outcome under control at time t depending on the strata at the same time; because of truncation of the outcome, wages are here defined only for the always-employed, EE, and the defiers, EN:

$$f(Y(C)_{i,t}|\mathbf{X}_i, G_{i,t} = EE, Z_i = C, \boldsymbol{\eta}^Y_{EE,C,t}) : N(\alpha^Y_{EE,t,C} + \mathbf{X}_i^T \boldsymbol{\beta}^Y_t; \sigma^2_{C,t})$$

$$f(Y(C)_{i,t}|\mathbf{X}_i, G_{i,t} = EN, Z_i = C, \boldsymbol{\eta}^Y_{EN,C,t}) : N(\alpha^Y_{EN,t,C} + \mathbf{X}_i^T \boldsymbol{\beta}^Y_t; \sigma^2_{C,t})$$

- the outcome under treatment at time t depending on the strata at the same time; because of truncation of the outcome, wages are here defined only for the always-employed, EE, and the compliers, NE:

  $$f(Y(T)_{i,t}|\mathbf{X}_i, G_{i,t} = EE, Z_i = T, \boldsymbol{\eta}^Y_{EE,T,t}) : N(\alpha^Y_{EE,t,T} + \mathbf{X}_i^T \boldsymbol{\beta}^Y_t; \sigma^2_{T,t})$$

  $$f(Y(T)_{i,t}|\mathbf{X}_i, G_{i,t} = NE, Z_i = T, \boldsymbol{\eta}^Y_{NE,T,t}) : N(\alpha^Y_{NE,t,T} + \mathbf{X}_i^T \boldsymbol{\beta}^Y_t; \sigma^2_{T,t})$$

- the response indicator under control at time t depending on the strata at the same time:

  $f(R(C)_{i,t}|\mathbf{X}_i, G_{i,t} = g, Z_i = C, \boldsymbol{\eta}^R_{gC,t})$: Logit$(\alpha^R_{t,C} + \mathbf{X}_i^T \boldsymbol{\beta}^R_t + \gamma^R_{NE,t,C} NE_{i,t} + \gamma^R_{EE,t,C} EE_{i,t} + \gamma^R_{NN,t,C} NN_{i,t})$,

  analogous formulation holds for the response indicators under treatment.

  The complicated structure of the resulted posterior distribution (2) can be adequately addressed by adopting a DA algorithm (Tanner and Wong (1987)), by exploiting the fact that with $(G_{i,18}, G_{i,36})$ known the likelihood loses its mixture structure.

4 Data

In order to evaluate the causal effect of language training programs on wages, we need information on pre-treatment individual characteristics and post-labour market outcomes, which are gathered by combining two rich datasets.

The first dataset is represented by administrative records derived from the global social security database in Luxembourg (Inspection Générale de la Sécurité Sociale (IGSS)), and collects social security forms of all workers employed in the country since 1980. These data allow us to follow workers trajectories from their first entrance in the labor market by personal identification number. It represents a rich reference source, given its detailed longitudinal information and the inclusion of natives, and immigrants. The quality of the data is very high. They are in fact used for calculating pensions in Luxembourg and regularly updated[4]. The second data source is a panel data on training programs collected by the Unemployment Agency (ADEM) in Luxembourg. The observation unit is represented by an "unemployment file", which corresponds to an unemployment spell. Any request by an individual for registration with ADEM consequently results in the opening of an "unemployment file", which is closed when the unemployed no longer checks-in at a meeting scheduled by the agency [5].

A rich set of information for the linked unemployed worker registered in ADEM is available from January 2007 to January 2012: age, education, gender, nationality, date of start of job, wage, number of hours worked, firm size, profession, sector of activity, as well as date of registration with ADEM, duration of registration in months, civil status, status previous to unemployment registration, type of job required by the unemployed, type of interventions/programs implemented by the agency, a score variable assessing the employability level of the unemployed worker, and partly driving training assignment. In particular, it is worth noting the inclusion of the score variable in the analysis, which will allow us to better identify the underlying assignment mechanism to alternative labour market measures, making our empirical strategy unique in this context.

Table 1 shows the sample size of the population of interest, by treatment and employment status (18 months after entering unemployment).

[4]The dataset is a matched employer-employee database.

[5]For example, because of finding a job, missing the meeting, or dropping out of the labor market.

Table 1: Sample size

		Z		
		Control	Treatment	
S	Employed	14986	325	15311
	Not employed	16721	318	17039
		31707	643	32350

5 Results

5.1 Design phase

Since the lack of balance in the pre-treatment characteristics between the treated and the control group can make any subsequent analysis imprecise, as well as sensitive to minor changes in the model specification for the outcomes, we aim to build a sample where the pre-treatment distributions among the two groups are well balanced[6].

We use matching on the estimated propensity score[7] to create a control sample, selected from the large reservoir of control units (31707) available in the data, in such a way that the pre-treatment variables distribution in the matched control group is similar to the pre-treatment variables distribution in the treated sample. More specifically, the best control match for each

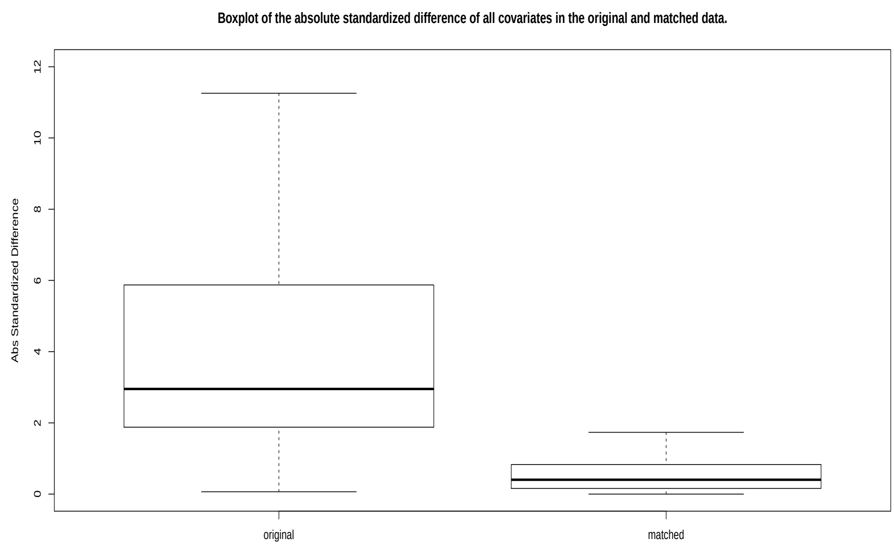
[6] This choice is also justified in light of the sensitivity analysis conducted by Bia et al. (2017) on similar data. They implemented a sensitivity analysis to account for unobserved confounding and found that the estimated results were robust to departures from uncounfoundedness assumptions.

[7] Let $p(X)$ be the probability of being assigned to the training given the set of covariates X: $p(X) = Pr(Z = 1|X = x) = E[Z|X = x]$. Rosenbaum and Rubin (1983) show that if the potential outcomes $Y(0), Y(1)$ are independent of treatment assignment conditional on X: $Y(0), Y(1) \perp Z|X$ (unconfoundedness assumption), they are also independent conditional on $p(X)$: $Y(0), Y(1) \perp Z|p(X)$.

treated unit is selected using the estimated propensity score[8] as a distance measure, that is, the control unit closest to the treated unit on the distance measure (nearest neighbor).

Figure 1 shows the absolute standardized difference of all covariates before and after matching. It is evident the great improvement in balancing the pre-treatment characteristics of the two groups when considering the selected individuals. Therefore, our analysis is performed on this subsample of units and the relative estimated results are reported in Table 2 and 3 of section 5.2.

Figure 1:



[8]A logistic regression model on the set of pre-treatment variables has been implemented to estimate the propensity score.

5.2 Preliminary results

In Table 2 we reported the effect of language training programs on the hourly wage and employment at 18 and 36 months after entering unemployment, respectively. The estimated effects on employment ($\hat{\pi}_{NE} - \hat{\pi}_{EN}$) are always positive and statistically significant, but higher in the first period (around 8% at 18 months and 3.3% at 36 months). The effect of foreign language programs on the wage for always-employees is slightly negative (-0.6) in the first period and closer to 0 (-0.14) in the second one, but never statistically significant.

From a policy point of view, these findings indicate that the language training programs have been successful in augmenting re-employment probabilities in both periods, but failed in providing unemployed with substantial human capital, with no increase in the wage offered to the trainees. This might be an incentive for ADEM to better design future policies implemented in the context of language trainings.

Of course, this part of the analysis is drawing results for those always employed, the only group of people for whom we can observe wages both under treatment and control and derive meaningful inferences. Nevertheless, inferences about the other strata can also provide interesting and additional insights about the intervention. Indeed, as already stressed in the first section of the paper, a key objective of our study is to investigating the behavior of defiant employees over time, whose wages can be not defined just because they would have a higher reservation wage under treatment. In other words, these individuals might be offered a job after the training, which they likely tend to refuse because they feel better equipped.

We investigate this hypothesis looking at the posterior probabilities of transitioning from a stratum at 18 months to another at 36 months after registering at ADEM (see Table 3). Specifically, we focus on the probability of being defiant employees (EN) at 18 months and becoming always employees (EE) at 36 months. This probability is equal to 0.195, which combined with the probability of being in the EN stratum in the first period, 0.497, reveals that the highest transition probability between 18 and 36 months is the one from defiant-employees to always-employees: defiant-employees at time $t = 18$ switch to the EE stratum with a proportion of almost 50%. This is in line with the labor economic theory and it is exactly what our study brings to evidence. Defiant-employees reasonably think the training course improve their job skills and so tend to wait more time before exiting the unemployment status in order to find a job better rewarded later on.

Table 2: Posterior means and standard devations

	at 18 months		at 36 months	
$\hat{\pi}_{EE}$	$.065$	(.00)	$.399$	(.04)
$\hat{\pi}_{NE}$	$.497$	(.04)	$.140$	(.03)
$\hat{\pi}_{EN}$	$\mathbf{.409}$	(.01)	$\mathbf{.107}$	(.00)
$\hat{\pi}_{NN}$	$.029$	(.01)	$.354$	(.02)
Est. effect on employment $\hat{\pi}_{NE} - \hat{\pi}_{EN}$	$\mathbf{.08}$	(.01)	$\mathbf{.033}$	(.00)
$Ave\widehat{Treated}EE(T)$	12.68	(.27)	15.43	(.34)
$Ave\widehat{Treated}EE(C)$	13.33	(.91)	15.57	(.35)
$Ave\widehat{Treated}NE(T)$	15.03	(.27)	16.22	(.57)
$Ave\widehat{Treated}EN(C)$	14.79	(.25)	15.52	(.93)
Est. effect on hourly wages for treated EE	$-\mathbf{.64}$	(1.95)	$-\mathbf{.14}$	(.46)

Table 3: Posterior means and standard devations for the joint probabilities

$\hat{\pi}_{NE.NE}$	.139 (.02)
$\hat{\pi}_{NE.NN}$	.212 (.03)
$\hat{\pi}_{NE.EE}$	.145 (.02)
$\hat{\pi}_{NN.NE}$	.000 (.00)
$\hat{\pi}_{NN.NN}$	.028 (.01)
$\hat{\pi}_{NN.EE}$	.002 (.00)
$\hat{\pi}_{EE.NE}$	.001 (.00)
$\hat{\pi}_{EE.NN}$	.007 (.00)
$\hat{\pi}_{EE.EE}$	.057 (.02)
$\hat{\pi}_{EN.EN}$	.107 (.00)
$\hat{\pi}_{EN.NN}$	.107 (.00)
$\hat{\pi}_{EN.EE}$	**.195** (.01)

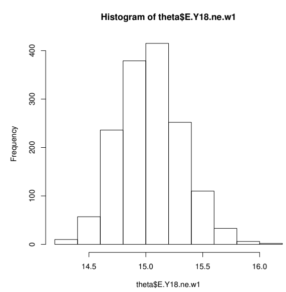
6 Conclusions

In this paper we analyze the short- and long-run effect of foreign language training programs on employment and wages measured over time, using administrative data on labour force in Luxembourg (IGSS-ADEM dataset). We use longitudinal information on these two outcomes at 18 and 36 months after entering unemployment and introduce a novel framework to simultaneously handle truncated wages due to unemployment, with incomplete observations not ignorable over time. Our model allows us to define important subpopulations of interest for policy making, with a focus on defiant-employees' behavior, and analyze the data more in detail than is possible via the standard selection models (as Heckman selection models), exploiting its longitudinal structure. More specifically, our findigns indicate that language trainings have been effective in increasing re-employment probabilities, but failed in providing unemployed people with substantial human capital, with no effect on the wages offered to the trainees. This might be an incentive for ADEM to design future policies, in the context of foreign language programs, better targeted to desired labor market outcomes.

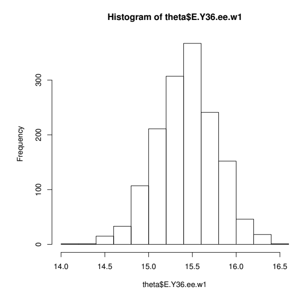
We then focus the analysis on defiant-employees and find that the highest transition probability between the two periods is the one of defiers at 18 months, who switch to the always-employees stratum at 36 months, with a proportion of almost 50%. This empirical evidence is in line with the labor economic theory, showing that defiers exposed to the training feel better equipped at the end of the program, hence increasing their reservation wage, and reasonably waiting more time before exiting the unemployment status in order to get a job better paid later on.

A Appendix

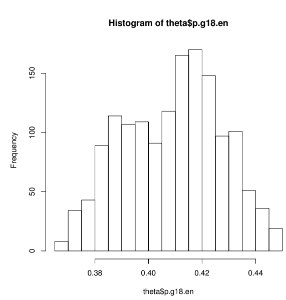
Figure 2: Histograms of the estimated wages (a, b) and strata probabilities (c, d)



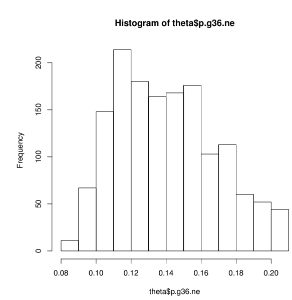
(a)



(b)



(c)



(d)

References

Bia, M., Flores-Lagunes, A., and Mercatanti, A. (2017). Evaluation of language training programs using principal stratification: The case of luxembourg. *Liser working paper, forthcoming*.

Brown, S. and Taylor, K. (2013). Reservation wages, expected wages and unemployment. *Economics Letters*, 119:276–279.

Frangakis, C. E. and Rubin, D. B. (2002). Principal stratification in causal inference. *Biometrics*, 58:21–29.

Frumento, P., Mealli, F., Pacini, B., and Rubin, D. B. (2012). Evaluating the effect of training on wages in the presence of noncompliance, nonemployment, and missing outcome data. *Journal of the American Statistical Association*, 107:450–466.

Imbens, G. and Rubin, D. B. (1997). Estimating outcome distributions for compliers in instrumental variables models. *The Review of Economic Studies*, 64:555–574.

Imbens, G. and Rubin, D. B. (2015). Causal inference for statistics, social, and biomedical sciences. *Cambridge University Press*.

Jones, S. (1988). The relationship between unemployment spells and reservation wages as a test of search theory. *Quarterly Journal of Economics*, 15:741–65.

Krueger, A. B. and Mueller, A. J. (2016). A contribution to the empirics of reservation wages. *American Economic Journal: Economic Policy*, 8:142–179.

Lancaster, T. and Chesher, A. (1983). An econometric analysis of reservation wages. *Econometrica*, 51:1661–76.

Little, R. J. and Rubin, D. B. (2002). Statistical analysis with missing data. *Wiley Series in Probability and Statistics*.

Mattei, A., Li, F., and Mealli, F. (2013). Exploiting multiple outcomes in bayesian principal stratification analysis with application to the evaluation of a job training program. *The Annals of Applied Statistics*, 7:2336–2360.

Mattei, A., Mealli, F., and Pacini, B. (2014). Identification of causal effects in the presence of nonignorable missing outcome values. *Biometrics*, 70:278–288.

Mealli, F. and Rubin, D. B. (2015). Clarifying missing at random and related definitions, and implications when coupled with exchangeability. *Biometrika*, 102:995–1000.

Mortensen, D. (1986). *Job Search and Labor Market Analysis*. O. Ashenfelter and R. Layard, Handbook of Labor Economics, North-Holland.

Rhosholm, M. and Svarer, M. (2008). The threat effect of active labour market programmes. *The Scandinavian Journal of Economics*, 110:385–401.

Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70:41–55.

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63:581–592.

Rubin, D. B. (1978). Bayesian inference for causal effects: The role of randomization. *The Annals of Statistics*, 6:34–58.

Rubin, D. B. (2000). The utility of counterfactuals for causal inference-discussion of causal inference without counterfactuals, by a. p. dawid. *Journal of the American Statistical Association*, 95:435–438.

Rubin, D. B. (2006). Causal inference through potential outcomes and principal stratification: Application to studies with censoring due to death. *Statistical Science*, 21:299–321.

Schafer, J. and Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Method*, 7:147–177.

Schmieder, F., Wachter, T. v., and Bender, S. (2016). The effect of unemployment benefits and nonemployment durations on wages. *American Economic Review*, 106:739–777.

Sørensen, K. L. (2017). Active labour market programmes and reservation wages: it is a hazard. *Applied Economics Letters*, 24:589–593.

Tanner, M. A. and Wong, W. H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, 82:528–540.

Van den Berg, G., Berghemann, A. H., and Caliendo, M. (2009). The effect of active labour market programs on not-yet treated unemployed individuals. *Journal of the European Economic Association*, 7:606–616.

Zhang, J., Mealli, F., and Rubin, D. B. (2008). *Evaluating the Effects of Job Training Programs onWages Through Principal Stratification*. D. Millimet, J. Smith, and E. Vytlacil, Oxford, UK: JAI Press.

Zhang, J. and Rubin, D. (2003). Estimation of causal effects via principal stratification when some outcomes are truncated by death. *Journal of Educational and Behavioral Statistics*, 28:353–368.

Zhang, J., Rubin, D., and Mealli, F. (2009). Likelihood-based analysis of causal effects of job-training programs using principal stratification. *Journal of the American Statistical Association*, 104:166–176.