Bia, Michela; Flores-Lagunes, Alfonso; Mercatanti, Andrea

Working Paper

Evaluation of Language Training Programs in Luxembourg using Principal Stratification

GLO Discussion Paper, No. 289

Provided in Cooperation with:
Global Labor Organization (GLO)

Evaluation of Language Training Programs in Luxembourg using Principal Stratification

Michela Bia *

Alfonso Flores-Lagunes †

Andrea Mercatanti ‡

November 28, 2018

*Evaluation Unit, LISER Luxembourg Institute of Socio-Economic Research, 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.

†Department of Economics and Center for Policy Research, Syracuse University, USA; IZA, Bonn, and Global Labor Organization (GLO). Email: afloresl@maxwell.syr.edu

‡Evaluation Unit, LISER Luxembourg Institute of Socio-Economic Research, Luxembourg, and Directorate General for Economics, Statistics and Research, Bank of Italy, Rome, Italy. Email: andrea.mercatanti@liser.lu.

ABSTRACT

In a world increasingly globalized, multiple language skills can create more employment opportunities. Several countries include language training programs in active labor market programs for the unemployed. We analyze the effects of a language training program on the re-employment probability and hourly wages of the unemployed simultaneously, using high-quality administrative data from Luxembourg. We address selection into training by exploiting the rich administrative information available, and account for the complication that wages are "truncated" by unemployment by adopting a principal stratification framework. Estimation is undertaken with a mixture model likelihood-based approach. To improve inference, we use the individual's hours worked as a secondary outcome and a stochastic dominance assumption. These two features considerably ameliorate the multimodality problem commonly encountered in mixture models. We also conduct sensitivity analysis to assess the unconfoundedness assumption employed. Our results strongly suggest a positive effect (of up to 12.7 percent) of the language training programs on the re-employment probability, but no effects on wages for those who are observed employed regardless of training participation. It appears that, in the context of an open and multilingual economy, language training improve employability but the language skills acquired are not sufficiently rewarded to be reflected in higher wages.

KEYWORDS: language training programs, policy evaluation, principal stratification, unconfoundedness, sensitivity analysis.

JEL CODES: C21, I38, J38

1 Introduction

Multiple language proficiency is becoming increasingly important for both developed and developing countries given the growing interconnection of nations within an increasingly globalized world. International communication represents a crucial pre-condition for promising international trade and financial investments. Moreover, within multilingual countries, multiple language skills significantly reduce information costs and help economic agents establish long run business relations. They also help firms build relationships with immigrant communities in the host country. These factors lead to an increasing demand for multiple language skills, which are not always met by skilled supply (Isphording 2014).

From an economic perspective, the market value of speaking a language is determined by, among other factors, the relative importance of a language in a given country (Isphording, 2014) which is itself a function of the language diversity within the country and its degree of international integration. It is also determined by the economic importance of commerce by immigrants in the host country (Lohman, 2011; Isphording and Otten, 2013). A range of studies document positive effects of language related skills on labor market outcomes such as earnings and employment (e.g., Dustmann and Fabbri, 2003; Williams, 2011; Ginsburgh and Prieto-Rodriguez, 2011; Isphording, 2013; Donado, 2017) that are also present throughout the earnings and occupational distributions (Ginsburgh and Prieto-Rodriguez, 2011; Isphording, 2014). The benefits of language skills have been found to potentially go beyond the labor market, such as the increased recognition among peers (Church and King, 1993; Selten and Pool, 1991).

As a result of the increasing importance of multiple language skills, several countries provide language training for the unemployed through their active labor market programs or ALMPs (e.g., McHugh and Challinor, 2011). These programs are funded and administered in a variety of ways, such as in a decentralized manner that delegates to states and community colleges as it is done in the U.S. (e.g., McHugh and Challinor, 2011); or by having a comprehensive federal-level strategy as in Germany (OECD, 2007). There is considerable heterogeneity in the structure of language training classes as well. A popular approach, typically targeted to recent immigrants, consists of delivering basic language skills. However, the market value of this type of training is debatable, likely due to a mismatch with the set of language skills needed by trainees in their typical occupations (McHugh and Challinor, 2011; Clausen et al., 2006). Language courses targeted to particular sectors of employment that are also designed with the demand for language skills in mind are believed to have a higher likelihood of boosting the labor market prospects of trainees (McHugh and Challinor, 2011). In the context of ALMPs in Switzerland, a multilingual country, Gerfin and Lechner (2002) found negative effects on employment from language courses. Against this backdrop, it is important to evaluate the existing language skills training programs for the unemployed in an effort to increase knowledge about relevant aspects that contribute to the improvement of participants' labor market prospects. We undertake such an analysis in the context of Luxembourg, a small open economy with a multilingual population and an important set of ALMPs.

The Employment Agency in Luxembourg (ADEM) is responsible for the country's ALMPs for the unemployed. ADEM delivers a wide range of training programs, among them language training programs. Language classes are offered to unemployed individuals to help improve job seekers' skills and better equip them for the labour market. Given the status of Luxembourg as a multilingual country, this type of programs is considered an effective instrument to tackle unemployment, especially among young people (European Social Fund in Luxembourg: Report of the European Commission, May 2015). However, to date there has not been an evaluation of this type of program. We formally evaluate the effect of attending ADEM's language training programs on the re-employment probability and on the hourly wage of re-employed individuals (who are re-employed irrespective of language training participation)

18 months after entering unemployment. We use administrative data from ADEM and from the Luxembourguish Global Social Security Database on Labour Force (IGSS). Our sample consists of 597 unemployed individuals who attended a language course (the treated group), for whom we evaluate the effects of the program by exploiting the information from a large reserve of unemployed individuals $(25,931)$ who did not participate in any type of ADEM training programs during the same period of time (the untreated group).

In our observational study, we need to deal with two main complications. The first is the selection of individuals into the language training programs. In the case of Luxembourg, the selection mechanism is a combination of individual willingness to take part on language classes, coupled with the administrative determination by a caseworker in ADEM that such training is a good choice.[1] To account for this selection, we assume unconfoundedness, which is justified on the basis of the rich administrative data we have available. The second complication is specific to the hourly wage outcome we consider: wages are only defined for those individuals who are employed 18 months after registering with ADEM (i.e, wages become "truncated" by unemployment). To tackle this complication we employ the framework of principal stratification (Frangakis and Rubin, 2002), which allows undertaking causal inference on individuals that would be employed irrespective of their participation in language training programs (a principal stratum).[2]

[1]See
http://www.adem.public.lu/en/marche-emploi-luxembourg/acteurs/adem/demandeurs-emploi/index.html;
http://www.adem.public.lu/en/demandeurs-demploi/sinscrire-a-ladem/pourquoi-sinscrire/index.html.
(Both accessed July 7, 2018)

[2]A different approach to deal with selection into employment consists of using exclusion restrictions for identification (see, e.g., Heckman, 1979 and Angrist and Krueger, 1999). However, finding variables that are related to employment but not related to hourly wages is typically challenging (Angrist and Krueger, 1999). Another approach to deal with selection into employment consists of using nonparametric bounds, as in Blundell et al. (2007) and Blanco et al. (2013).

Under principal stratification, the population is classified into latent principal strata based on the four potential values of an intermediate variable (employment) under each of the treatment arms. Within principal strata, comparisons of units under different treatment arms (possibly conditional on covariates) yield valid causal effects. As a result, under principal stratification, interest typically lies in estimating causal effects local to a particular principal stratum–in this case the stratum of individuals that would be employed irrespective of their participation in language training programs. Principal stratification has roots in causal models with instrumental variables (Imbens and Angrist, 1994; Angrist, Imbens and Rubin, 1996). Zhang et al. (2009) and Frumento et al. (2012) employed principal stratification to deal with the problem of selection into employment when considering wages as an outcome in the evaluation of a randomized training program implemented in the U.S. Given the randomized nature of the treatment in their context, they did not have to deal with the issue of selection into the treatment.

Our approach can be seen as an extension of the principal stratification approach to analyze effects on wages by Zhang et al. (2009) and Frumento et al. (2012) to the setting of an observational study. We deal with the non-random selection of unemployed individuals into language training by allowing for the probability of treatment assignment to depend on a rich set of observable individual characteristics that control for selection. Also, contrary to Zhang et al. (2009) and Frumento et al. (2012), who employ a direct likelihood approach, we employ a traditional likelihood approach (McLachlan and Peel, 2000, ch. 2 and 3) that allows the computation of standard errors for the estimated parameters. The resulting likelihood function from our model presents multimodality: a high number of local maxima that makes inference challenging. However, we demonstrate that this likelihood function can be largely regularized—and thus the number of local maxima reduced—by introducing a secondary outcome and a stochastic dominance restriction. The use of a secondary outcome had been advocated by Mattei et al. (2013) and Mercatanti et al. (2015) to sharpen inference both in a Bayesian and likelihood frameworks within the principal stratification approach, while a similar stochastic dominance restriction to the one used here was employed by Zhang et al. (2009) in a similar empirical setting. A final important practical aspect we consider in our observational setting is the implementation of a sensitivity analysis in the spirit of Rosenbaum (2002) to assess the robustness of our inference to unobserved factors that may impact the selection into language training.

This paper has four main contributions. First, we contribute to the growing literature on the labour market effects of language skills (e.g., Dustmann and Fabbri, 2003; Williams, 2011; Ginsburgh and Prieto-Rodriguez, 2011; Isphording, 2014; Donado, 2017). We do this by formally evaluating the labour market benefits of language training for the unemployed in Luxembourg, a multilingual country with a sizable proportion of immigrants. Second, we advance the empirical evidence on the effectiveness of training programs in Luxembourg. Only a few studies exist on the effectiveness of labour market policies in the country (e.g., Brosius and Zanardelli, 2012), where they report a positive effect of the bundle of training programs in ADEM's ALMPs on post-training employment in the short-term, but reduced effectiveness in the long-term. We provide evidence of the effectiveness of an important component of the bundle of training programs offered by ADEM: language training. Third, methodologically, we provide guidance on how to conduct a principal stratification analysis when both selection into treatment and truncation by unemployment (or another relevant intermediate variable) are considered, which is a frequent occurrence in the evaluation of public programs. Moreover, we demonstrate how the contemporary adoption of a secondary outcome and a stochastic dominance restriction help to overcome situations in which the regularity of the likelihood function is broken. Fourth, this study also illustrates how to conduct an analysis of sensitivity to the presence of unobserved factors that influence both the assignment into treatment and the outcomes of interest.

The rest of the paper is organized as follows. Section 2 provides background on Luxembourg and its ALMPs. Section 3 introduces the causal model, the underlying assumptions for identification of the causal effects of interest, and the likelihood function. Section 4 describes the administrative data from IGSS and ADEM in Luxembourg and provides descriptive statistics. The main results and their implications are also presented in this section. Section 5 undertakes a sensitivity analysis of our main results by accounting for possible impacts of unobservable variables on the treatment assignment. Section 6 concludes.

2 Motivation and Background

2.1 Luxembourg

Luxembourg is situated in Western Europe, it is landlocked, and borders with Belgium, Germany, and France. Its strategic geographical location has shaped the country as a multilingual and multicultural marketplace, unique all over Europe. Luxembourg has just over half a million inhabitants and it is a popular destination among expatriates, with around 45% of residents and 65% of the working population being foreign citizens (Statec, 2014a; Statec, 2014b). The share of expatriates in Luxembourg has more than doubled over the last 25 years with a large wave of Italian immigrants in the first half of the 1960s, followed by a relatively recent immigrantion wave coming mainly from Portugal. Portuguese expatriates became the largest foreign community in the country (Statec, 2012). In addition, the share of foreign nationals from neighboring countries has also been increasing over the last decades: from 1961 to 2011 the French population increased from 1.6% to 6.7%, the Belgian population from 1.7% to 3.3%, and the German population, more stable, from around 2.2% to 2.4% (Statec, 2013).

Luxembourg's culture is historically a combination of Romanic and Germanic philosophy and institutions. It is currently a trilingual country, with Luxembourgish, German, and French designated as official languages. Indeed, the different schooling levels are taught in the three different languages, with pre-school taught in Luxembourguish, elementary in German, and secondary in French. Multilingualism is, of course, one of the country's strengths in the face of an increasingly internationally integrated world. However, this also requires ad-hoc education and training programs, as well as efficient labour market integration policies. Noteworthy, a public agency for integration (Office Luxembourgeois de l'Accueil et de l'Intgration) provides immigrants with information on training in the official languages and recognition of foreign diploma and secondary education, while ADEM and the education ministry collaborate in providing alternative training courses for job seekers, with a focus on language courses primarily for foreigners. In this context, the experience of Luxembourg pertaining to active labour market programs for the unemployed is relevant to small open economies and to countries experiencing proportionately large migration inflows.

2.2 Active labour Market Programs in Luxembourg

Luxembourg mirrors the objectives and challenges of several European countries, such as ensuring access and progression in economic opportunities to the general population and to the unemployed in particular, irrespective of their linguistic and socio-economic conditions.[3] A variety of training programs have been introduced over the last decades, both in Europe and North America, to improve immigrants' employment opportunities through language acquisition. Among the goals of these programs—including the existing programs in Luxembourg—is to encourage immigrants to enter the country's formal labour market and to help them move into better-paid jobs. However, cost-effective language courses are a challenge, since policy

[3] See http://www.adem.public.lu/en/demandeurs-demploi/sinscrire-a-ladem/personnes-concernees/index.html (accessed July 7, 2018).

makers have to design interventions tailored to immigrants' employment needs, their cultural background, and family conditions (McHugh and Challinor, 2011).

The language courses offered by ADEM's training consist mainly of Luxembourgish, German, French, and English, with an average duration of 5 months. They are provided either alone or in combination with a variety of other complementary ALMP schemes.[4] We focus here on unemployed individuals who exclusively enrolled in language training programs. The language courses by ADEM are certified by the "Institut National des Langues" (http://www.inll.lu/en/), located in the city of Luxembourg. Special exams are given at the end of each course in order to test the level of proficiency in listening, reading, writing and speaking achieved by the unemployed in a given language.[5]

ADEM implements a "personalized assistance" model tailored to the needs of the individual unemployed. After an initial interview with a professional counselor, the unemployed individual is referred to an assistance scheme that best corresponds to his or her own profile. The aim of ADEM is to remove obstacles preventing job seekers from entering the labour market.[6] If deemed necessary, case-workers assign the unemployed to a given training or ALMP taking into account all the individual's information, such as educational level, health and psychological status, job expertise, and preferences in terms of job sought. In case of perceived communication barriers related to language, training language courses are among the first suggested and offered to the unemployed. Notably, the individual information available

[4]For additional details see Brosius and Zanardelli, "Evaluation de lefficacité des mesures de formation destinées aux chômeurs", Report provided to the Ministry of Labour, Employment and the Social and Solidarity Economy of Luxembourg in 2012 and the review made by Patrick Till in 2015 for the European Employment Policy Observatory (EEPO), "Review Spring 2015: Upskilling unemployed adults".

[5]See, for example, http://www.inll.lu/en/certifications-nationales-et-internationales/apercu/ (accessed July 7, 2018).

[6]See http://www.adem.public.lu/en/demandeurs-demploi/sinscrire-a-ladem/encadrement/index.html (accessed July 7, 2018).

to ADEM's case-workers is summarized in a score related to the individual's employability level, which is a variable available in our administrative data.

Few studies exist on the effectiveness of labour market policies in Luxembourg (e.g., Brosius and Zanardelli, 2012; OECD, 2010, 2012). At the same time, the OECD (2010, 2012) reports point out that " ...job prospects amongst unemployed and cost effectiveness would benefit from a better design of labour market programs in Luxembourg". In this context, a contribution of our study is to increase the amount of empirical evidence on the effectiveness of ALMPs in Luxembourg by focusing on language training programs that are key for a multilingual country with a sizable number of immigrants. Evaluating the effectiveness of language training on subsequent labour market outcomes is an important first step in assessing ways to improve them and identifying best practices.

3 Methodology

3.1 General Framework and Notation

We adopt the potential outcomes framework or Rubin Causal Model (RCM) to define causal effects (Rubin, 1974, 1978). Consider a sample of N units. For each unit i let Z_i be a binary treatment variable, equal to one if the unemployed individual receives language training, and 0 if he does not receive language training. Let $Y_i(Z_i)$ denote the potential outcomes for individual i, namely the the potential value of the outcome under each of the two possible treatment assignments. In our context, Y_i represents the hourly wage 18 months after entering ADEM, one of our two outcomes of interest. In addition, let X_i be a vector of pre-treatment characteristics. Let S_i be a binary post-treatment variable, equal to 1 if subject i is employed 18 months after registration at ADEM, 0 otherwise. In our context, S_i is the second outcome of interest that determines the observability of Y_i and that is likely affected by Z_i. We denote the potential values of this variable as a function of the treatment as $S_i(Z_i)$. The goal is to identify and estimate the causal effect of Z_i on both outcomes of interest, S_i and Y_i.

To identify the effect of Z_i on Y_i, two problems have to be tackled. The first one is the self-selection of the unemployed into the treatment. Namely, how is it that the units we observe with $Z_i = 1$ came to receive language training? The second problem is "selection into employment", that is, the wages of individuals in the sample are only observed conditional on them being employed. The second issue relates the two outcomes of interest (Y_i and S_i). Note that, to identify the effect of Z_i on S_i, only the first of the two problems arises. To address the first identification problem, we assume that assignment to the treatment is strongly ignorable (Rosenbaum and Rubin, 1983a), which we formalize in the next subsection. To address the second identification problem, we adopt the principal stratification framework (Frangakis and Rubin, 2002). The population is partitioned into four latent groups based on the values of the vector $\{S_i(1), S_i(0)\}$, called principal strata:

EE: subjects who would be employed regardless of treatment assignment:

$$EE = \{i : S_i(1) = 1, S_i(0) = 1\}.$$

EN: subjects who would be employed under treatment, but not employed under control:

$$EN = \{i : S_i(1) = 1, S_i(0) = 0\}.$$

NE: subjects who would not be employed under treatment but employed under control:

$$NE = \{i : S_i(1) = 0, S_i(0) = 1\}.$$

NN: subjects who would not be employed regardless of treatment assignment:

$$NN = \{i : S_i(1) = 0, S_i(0) = 0\}.$$

Denote the proportion of individuals in the population belonging to each one of these latent groups as π_{EE}, π_{EN}, π_{NE}, and π_{NN}, respectively. The importance of partitioning the population into principal strata is that, within strata, the comparisons of potential outcomes can be given causal interpretation (Frangakis and Rubin, 2002). In other words, even though S_i may be affected by the treatment, by focusing on units that share the same potential values $\{S_i(1), S_i(0)\}$, causal effects of the treatment on Y_i can be identified. A simplistic alternative analysis that does not account for the problem of selection into employment would use only the individuals for whom wages are observed, namely those i for whom $S_i(Z_i) = 1$ (those employed). However, this approach would lead to results that lack causal interpretation. In fact, the units i such that $\{i : S_i(1) = 1\}$ are a mixture of EE and EN, while those such that $\{i : S_i(0) = 1\}$ are a mixture of EE and NE. Thus, this alternative analysis is at odds with the basic requirement that causal effects are defined as a comparison of potential quantities on a common set of units (Frangakis and Rubin, 2002). In the balance of this section, we define the causal effects of interest and discuss their identification, followed by the statistical model to be employed in their estimation.

3.2 Causal Effects of Interest and their Identification

The first parameter we are interested in is the (causal) average treatment effect (ATE) on re-employment 18 months after registration with ADEM, that is, $E[S_i(1) - S_i(0)]$. Using the notation introduced in the last section, it is straightforward to see that this effect can be defined as the difference in the following two population proportions: $\pi_{EN} - \pi_{NE}$. As for the causal effect of the treatment on the hourly wage 18 months after registration with ADEM, recall that the hourly wage is only defined conditional on $S_i = 1$. Therefore, we concentrate on the principal average causal effect (PACE) for the stratum of individuals that would be employed regardless of treatment assignment: $E[Y_i(1) - Y_i(0)|EE]$. This is a commonly estimated parameter in the literature (e.g., Zhang et al., 2009; Lee, 2009; Blanco et al., 2013), since this stratum is the only one for which the wage is observed under both treatment arms.[7]

To identify the two causal effects of interest, we adopt the following assumptions.[8]

Assumption 1 (Unconfoundedness): $Z_i \perp \{Y_i(0), Y_i(1), S_i(0), S_i(1)\}|X_i$.

Assumption 2 (Overlap condition): $0 < \Pr(Z_i = 1|X_i) < 1$, for all i.

Assumption 1 states that, conditional on observable variables X_i, the treatment is independent of both pairs of potential outcomes. This assumption, although widely used in the literature (e.g., Heckman, LaLonde and Smith, 1999; Imbens, 2003), is strong given that it rules out any unobserved confounders that are related to each of the potential outcomes and to the probability of receiving the treatment (after conditioning on X_i). Nevertheless, we believe that the combination of access to rich administrative data and the institutional features of the assignment of unemployed individuals into language training programs makes unconfoundedness a tenable assumption. We will further discuss its plausibility in Section 4 and conduct a sensitivity analysis to departures from it in Section 5. Assumption 2 states that the probability of undergoing the treatment (conditional on X_i) is bounded away from zero or one. In practice, this assumption requires that there are individuals with the same values of X_i who are observed in each of the two treatment arms.

In addition to the assumptions above, we impose the following stochastic dominance assumption in some of our models:

[7] This parameter is also known as the survivor average causal effect or SACE (Zhang et al., 2009) and as the (local) average treatment effect for the always-employed or ATE_{EE} (Blanco et al., 2013).

[8] In addition to the assumptions below, the "stable unit treatment value assumption" (SUTVA) is also adopted (Rubin, 1980). SUTVA rules out interference among individuals and any hidden versions of the treatment under consideration.

Assumption 3 (Stochastic Dominance): *For any real number t,* $P(Y_{EE}(1) \leq t) \leq P(Y_{EN}(1) \leq t)$ *and* $P(Y_{EE}(0) \leq t) \leq P(Y_{NE}(0) \leq t)$.

This assumption states that the wage distribution of the EE when trained stochastically dominates the wage distribution of the EN when trained, and that the wage distribution of the EE when not trained stochastically dominates the wage distribution of the NE when not trained. In other words, the assumption formalizes the notion that the EE group likely possesses characteristics that allows it to have higher or comparable wage-earning potential relative to both the EN and NE groups. Similar stochastic dominance assumptions were employed by Zhang et al. (2008), Zhang et al. (2009), and Blanco et al. (2013) in the context of estimating or constructing bounds on similar treatment effects. Here, however, we employ this assumption as a restriction on the likelihood function of our model that helps regularizing it and improves inference, as explained later.

To identify the parameters of interest, we combine the three assumptions above with a parametric model and employ mixture model analysis in the spirit of Zhang et al. (2009). In general, identification follows from combining a proposed parametric model for the potential outcomes with one for the principal strata membership, leading to a mixture model that is estimated via maximum likelihood (see Zhang et al., 2009). An important difference with Zhang et al. (2009), however, is that the covariates in X_i not only improve precision, but they also play the crucial role of controlling for selection into the language training program (following Assumption 1 and Assumption 2).

3.3 Estimation

To estimate the causal effects of interest, we construct a likelihood function based on the models for the potential outcomes and the principal strata membership. The two causal effects of interest are simultaneously estimated along with other parameters of the models. This approach requires the prediction of the individuals' missing membership to the principal strata. The membership is unknown since one potential value of $S_i(Z_i)$ is missing as $S_i^{\text{mis}} = S_i(z) : z \neq Z_i^{\text{obs}}$, where the superscripts mis and obs denote the missing and observed values of a variable, respectively. Similarly, each individual in the sample has a missing potential outcome as determined by $Y_i^{\text{mis}} = Y_i(z) : z \neq Z_i^{\text{obs}}$. Because we condition the analysis on the empirical distribution of the pre-treatment variables, $Pr(X_i)$ does not need to be modelled. Additionally, Assumptions 1 and 2 (unconfoundedness and overlap, respectively) imply that we can ignore the assignment mechanism $Pr(Z_i|Xi)$. Thus we focus on the distribution of the potential quantities $Y_i(Z_i)$ and $S_i(Z_i)$ given the pre-treatment variables which, by integration over the missing quantities, yields the following likelihood:

$$
\begin{aligned}
\mathcal{L}\left(\boldsymbol{\theta}^S, \boldsymbol{\theta}^Y; \mathbf{Z}^{\text{obs}}, \mathbf{S}^{\text{obs}}, \mathbf{Y}^{\text{obs}}, \mathbf{X}\right) = \prod_i \Bigg[\int \int & Pr\Big(S_i(0), S_i(1)|X_i; \theta^S\Big) \cdot \\
\cdot Pr\Big(Y_i(0), Y_i(1)|S_i(0), S_i(1), X_i; \theta^Y\Big) & dY_i^{\text{mis}} dS_i^{\text{mis}} \Bigg]
\end{aligned}
$$

where $\boldsymbol{\theta}^S$ and $\boldsymbol{\theta}^Y$ collect the parameters representing the proportions of individuals in the population in each one of the principal strata and the parameters of the conditional distribution of the potential outcomes of Y given principal strata membership, respectively.

More specifically, we employ the following logistic model for the principal strata membership:

$$P(G_i = g) = \pi_{g:i} = \frac{\exp(\mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta}_g)}{\sum_{g'}\exp(\mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta}_{g'})}, g \in \{EE, EN, NE, NN\}$$

where $G_i = g$ denotes membership to principal strata $g \in \{EE, EN, NE, NN\}$, and $\boldsymbol{\beta}_g$ are the model's parameters. We will choose, without loss of generality, the NN stratum as the baseline group (i.e., $\boldsymbol{\beta}_{NN} = 0$). The potential outcomes model for wages is specified as log-normal and allowed to vary by treatment status:

$$\begin{aligned}
\text{if } G_i = EE \quad &, \quad \log[Y_i(1)] \sim N(\mathbf{X}_i^1\boldsymbol{\eta}_{EE,1}, \sigma^2_{EE,1}) \\
&\quad\;\; \log[Y_i(0)] \sim N(\mathbf{X}_i^0\boldsymbol{\eta}_{EE,0}, \sigma^2_{EE,0}) \\
\text{if } G_i = EN \quad &, \quad \log[Y_i(1)] \sim N(\mathbf{X}_i^1\boldsymbol{\eta}_{EN,1}, \sigma^2_{EN,1}) \\
\text{if } G_i = NE \quad &, \quad \log[Y_i(0)] \sim N(\mathbf{X}_i^0\boldsymbol{\eta}_{NE,0}, \sigma^2_{NE,0})
\end{aligned}$$

After inserting the above models into the general formulation of the likelihood function, it can be factored into two mixtures of normal distributions and two sums of strata probabilities as follows:

$$\begin{aligned}
&\mathcal{L}(\boldsymbol{\theta}|\mathbf{Z}, \mathbf{S}^{obs}, \mathbf{Y}^{obs}, \mathbf{X}) \propto \\
&\quad \prod_{i\in(Z_i=1,S_i^{obs}=1)} [\pi_{EE:i}N_i(\mathbf{X}_i^1\boldsymbol{\eta}_{EE,1}, \sigma^2_{EE,1}) + \pi_{EN:i}N_i(\mathbf{X}_i^1\boldsymbol{\eta}_{EN,1}, \sigma^2_{EN,1})] \times \\
&\quad \prod_{i\in(Z_i=1,S_i^{obs}=0)} (\pi_{NE:i} + \pi_{NN:i}) \times \\
&\quad \prod_{i\in(Z_i=0,S_i^{obs}=1)} [\pi_{EE:i}N_i(\mathbf{X}_i^0\boldsymbol{\eta}_{EE,0}, \sigma^2_{EE,0}) + \pi_{NE:i}N_i(\mathbf{X}_i^0\boldsymbol{\eta}_{NE,0}, \sigma^2_{NE,0})] \times \\
&\quad \prod_{i\in(Z_i=0,S_i^{obs}=0)} (\pi_{EN:i} + \pi_{NN:i}) \qquad (1)
\end{aligned}$$

The maximization of the above likelihood function is undertaken using the EM (expectation-maximization) algorithm (Dempster et al., 1977). In the expectation step, the unobserved principal strata are replaced by their expectations given the data and current estimates of the potential outcomes model parameters. Then, in the maximization step, the likelihood function, conditional on the expected principal strata, is maximized. Upon convergence of the algorithm, all parameters of the principal strata and potential outcomes models are obtained, and from them the causal effects of interest are calculated. The standard errors of all estimated parameters are obtained by relying on their asymptotic distribution using the outer product of gradients (McLachlan and Peel, 2000). We note that our estimation approach departs from Zhang et al. (2009) in that they use a direct likelihood approach that does not allow them to calculate standard errors but where alternative nested models can be compared using values of the log-likelihood function.

In practice, the likelihood function resulting from mixture models with normal components, like ours, presents a high number of local maxima (i.e., multimodality) that makes inference challenging (McLachlan and Peel, 2000, Ch. 2 and 3). This non-regularity feature of the likelihood function arises because the regularity conditions for the maximum likelihood (ML) estimator (Lehmann and Casella, 1998, Ch. 6) do not hold globally, but locally. Consequently, given that the ML estimator is not guaranteed to be the efficient likelihood estimator, the issue arises as to how to detect the local ML point that corresponds to this efficient estimator. Some proposals are available in the literature, such as selecting the one that is closest to a moments estimator (Lehmann and Casella, 1998), imposing suitable constraints on the variances of the mixture components (Hathaway, 1985; Aitkin and Rubin, 1985), or penalizing the likelihood function (Ciuperca et al., 2003). Those proposals, however, have been implemented in the context of considerably simpler mixture models with few parameters (e.g., Mercatanti, 2013). The adoption of the above proposals in our model would considerably increase the computational burden. For this reason, we propose for a different approach.

To help regularize the likelihood function and ameliorate the multimodality problem, we employ a secondary outcome and also impose the restriction of stochastic dominance (Assumption 3). The resulting regularized likelihood function improves inference. Recent contributions in the causal and mixtures literature (Mattei et al., 2013; Mealli and Pacini, 2013; Mercatanti et al., 2015) show that the inclusion of a secondary outcome can greatly improve the inference for the primary outcome by providing extra information to predict the mixture membership and disentangle the mixtures. Recall that the primary outcome (Y) is the hourly wage 18 months after entering ADEM, which we denote as Y_1 hereafter to introduce a secondary outcome that will be denoted as Y_2. In general, a good choice for a secondary outcome is a variable that is highly correlated with the primary outcome (Mealli and Pacini, 2013). For this reason, we choose the number of hours worked as a secondary outcome (Y_2).[9] It has been documented in the labour economics literature that there is a strong correlation between hourly wages and the number of hours worked (e.g., Kuhn and Lozano, 2008). Including the secondary outcome to improve inference, the potential outcomes model (in more compact notation) becomes:

$$\text{if } G_i = g, (\log[Y_{1,i}(Z)], Y_{2,i}) \sim N(\mathbf{X}_i^{\mathrm{T}}\mathbf{H}_{g,z}, \mathbf{\Sigma}_{g,z})$$

where

$$\begin{aligned}
\mathbf{H}_{g,z} &= (\boldsymbol{\eta}_{1,g,z}, \boldsymbol{\eta}_{2,g,z})^{\mathrm{T}}, \\
\mathbf{\Sigma}_{g,z} &= \begin{pmatrix} \sigma^2_{1,g,z} & \sigma_{1,2,g,z} \\ \sigma_{1,2,g,z} & \sigma^2_{2,g,z} \end{pmatrix}.
\end{aligned}$$

The expanded outcome model is then inserted into the general formulation of the likelihood

[9] The number of hours worked are collected on a monthly basis and observed only for re-employed individuals 18 months after registering at ADEM.

function, along with the model for the principal strata, and then maximized using the EM algorithm.

4 Evaluation of Language Training Programs in Luxembourg

4.1 The data

To evaluate the causal effects of the language training programs in Luxembourg on re-employment and on wages, we combine two rich administrative datasets. The richness of the data, in particular in the availability of relevant pre-treatment individual characteristics, is instrumental in arguing the plausibility of our identifying assumptions.

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 the trajectory of workers from their first entrance in the labour market using their personal identification number. It represents a rich reference source, given its detailed longitudinal information and the inclusion of natives and immigrants. The data is regularly updated and its quality is very high, as it is officially used for calculating pensions in Luxembourg. The second source is a longitudinal data set 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 individual registration with ADEM results in the opening of an "unemployment file", which eventually is closed when the unemployed individual no longer checks-in at meetings scheduled by the agency because of, for example, having found a job or dropped out of the labour market. Information from the two data sources above is linked using the individual's personal identification number. We focus on unemployed individuals that registered with ADEM from January 2007 to

October 2011, who are linked to their administrative records in IGSS.

A rich set of information is available after the linkage: age, gender, education, civil status, number of children, prior language skills, health and psychological status, and nationality. Available is also information on the last job and the new job (if employed), such as starting date, wage, number of hours worked, firm size, profession, and sector of activity. There is also information related to the unemployment spell, such as the date of registration with ADEM, duration of registration in months, civil status previous to unemployment registration, type of job desired by the unemployed individual, type of interventions/programs implemented by the agency, and a score variable assessing the employability level of the unemployed worker, which is relevant for assignment to alternative labour market measures, such as language training. In sum, we have access to most variables that prior literature on the evaluation of ALMPs has identified as important in determining selection into training programs (e.g., Lechner and Wunsch, 2013).

Table 1 shows the sample sizes by language training participation status (the treatment) and by employment 18 months after registering with ADEM, which is one of our outcomes and the variable that determines observability of the wages (our second outcome). As can be seen from the table, our data contains 597 unemployed individuals who participated in a language training program, while there is a large pool of $25,931$ unemployed individuals who did not participate in any type of ADEM training programs during the same period of time. The table also shows that 316 of the unemployed who participated in a language training program are employed, which represents a 53% employment rate that is higher relative to the employment rate of those who did not participate in any training program (51%). This difference in employment rates represents an unadjusted effect that likely lacks causal interpretation since participation in language training programs is not determined at random.

Table 1: Sample sizes by language training participation and employment

		Language Training (Z)		
		No	Yes	
S	Employed	13222	316	13538
	Not employed	12709	281	12990
		25931	597	26528

Table 2 and 3 present summary statistics for selected variables in our sample. Table 2 shows that about 53% of our sample consists of males, and about 49% of individuals are married. In terms of education, 47% of individuals list as primary their highest level of education, 37% of them as secondary, and 18% as graduate. 23% are Luxembourg natives, 28% are Portuguese natives, while 13% are from neighboring France, Belgium, or Germany, 8% of individuals are from other European Union (EU) countries and 10% are from outside the EU. Only about 17% of individuals hold a valid driver's license, which is consistent with the disadvantaged nature of this sample. About 23% of individuals in our sample do not speak any Luxembourgish or German, and about 85% and 77% are fluent in Portuguese and Italian, respectively.

Table 2: Summary Statistics – Individual characteristics

		Mean	SD	N
Gender	Male	0.526	0.499	26,528
Age		36.894	11.552	26,528
Education	Primary	0.467	0.498	26,528
	Secondary	0.356	0.478	26,528
	Graduate	0.175	0.380	26,528
Nationality	France-Belgium-Germany	0.133	0.339	26,528
	Lux	.226	0.418	26,528
	Portuguese	0.278	0.448	26,528
	Other EU	0.083	0.277	26,528
	OtherNoEU	0.103	0.304	26,528
	Not available	0.174	0.379	26,528
Civil Status	Married	0.488	0.499	26,528
	Single Divorced Widowed	0.507	0.499	26,528
	Not available	0.004	0.064	26,528
Number of children		0.75	1.156	26,528
Driver's license		0.171	0.376	26,528
Language skills	Lux: none	0.116	0.321	26,528
	Lux: basic-medium	0.022	0.148	26,528
	Lux: good	0.860	0.346	26,528
	French: none	0.018	0.135	26,528
	French: basic-medium	0.031	0.174	26,528
	French: good	0.949	0.218	26,528
	German: none	0.111	0.314	26,528
	German: basic-medium	0.019	0.136	26,528
	German: good	0.869	0.336	26,528
	Portoguese: none	0.146	0.353	26,528
	Portoguese: basic-medium	0.006	0.080	26,528
	Portoguese: good	0.846	0.360	26,528
	Italian: none	0.212	0.409	26,528
	Italian: basic-medium	0.013	0.115	26,528
	Italian: good	0.773	0.418	26,528
Informatics skills	none	0.243	0.492	26,528
	basic-medium	0.003	0.056	26,528
	good	0.753	0.431	26,528

Data source: IGSS-ADEM data 2007–2011.

Table 3: Summary Statistics – Job characteristics

		Mean	SD	N
N. of months employed before (last 6 months)		3.901	2.491	19,315
N. of months employed before (last 12 months)		6.858	4.998	19,565
Professional Status	Blue collar worker	0.388	1.201	26,528
	White collar worker	0.384	0.486	26,528
	Public Employee	0.110	0.313	26,528
	Self-employed	0.001	0.037	26,528
	Independent-intellectual work	0.011	0.104	26,528
	Employed in agriculture	0.002	0.047	26,528
	Other	0.000	0.010	26,528
Sector	Agirculture	0.004	0.067	26,528
	Extractive Industries	0.000	0.010	26,528
	Manifacturing	0.033	0.180	26,528
	Electricity-Gas Supply	0.000	0.026	26,528
	Construction	0.086	0.280	26,528
	Commerce	0.086	0.281	26,528
	Hotels and Restaurants	0.09	0.289	26,528
	Transports	0.028	0.166	26,528
	Financial Sector	0.035	0.183	26,528
	Real Estate	0.151	0.350	26,528
	Public Administration	0.017	0.132	26,528
	Education	0.002	0.052	26,528
	Health	0.030	0.172	26,528
	Social Services	0.019	0.139	26,528
	Domestic Services	0.016	0.128	26,528
	Extra-activities	0.015	0.038	26,528
	Not available	0.391	0.487	26,528
Job sought	Liberal Arts-Technicians	0.114	0.317	26,528
	Directors-Managers	0.022	0.149	26,528
	Office Employees	0.160	0.366	26,528
	Sales-Person	0.09	0.294	26,528
	Agriculture-forest-worker, miners	0.011	0.108	26,528
	Worker in transportation-communication	0.035	0.185	26,528
	Craftman-manual worker	0.283	0.450	26,528
	Hotels, restaurants	0.087	0.282	26,528
	Other services	0.140	0.347	26,528
	No preference	0.047	0.2012	26,528
Employability Level	Score A - no intervention	0.077	0.266	26,528
	Score B - short-term interventn	0.196	0.397	26,528
	Score C - medium-term interventn	0.287	0.452	26,528
	Score D - medterm w/ social asst	0.09	0.262	26,528
	Score E - long-term intervention	0.039	0.195	26,528
	To be determined	0.304	0.460	26,528
N. of months prior to taking training		3.38	4.669	26,528

Data source: IGSS-ADEM data 2007 2011

The first two rows of Table 3 speak to the labour market attachment of individuals in our sample. For instance, they have worked, on average, 3.90 and 6.85 months out of the last 6 and 12 months, respectively. The table also gives a picture of the distribution of sectors in which individuals in our sample held their last job: about 15% of them have worked in the real estate sector, 9% of them have worked in hotels and restaurants, about 9% of them in commerce, and 9% of them in construction. During the interview with the case-worker, they have also reported their preferences in terms of job sought. Among the favourite categories we find: manual work (28%), followed by office employee job (16%). Finally, looking at the employability level—the score variable relevant in determining selection of the unemployed into a given training program—about 20% need short-term interventions against about 30% needing medium-term interventions. This variable's category of "to be determined" contains unemployed individuals for whom the case-worker at ADEM chose to delay assigning a value. Typically, this assignment is done at a later meeting of the individual with ADEM, but unfortunately such subsequent determination is not currently available to us. Lastly, the average number of months prior to taking training in the sample is 3.38.[10]

4.2 Results

We start by estimating the parameters of the model without the secondary outcome and without imposing the stochastic dominance assumption. We refer to this model as the unrestricted model, given by the likelihood in (1). Since we find evidence that the likelihood function of the unrestricted model presents several local solutions (multimodality), we move on to include a secondary outcome. Subsequently, we consider a model that includes a secondary outcome

[10]This control variable is assigned to the unemployed that do not take training using the procedure in Lechner (1999) that consists of randomly drawing training starting dates for them from the empirical distribution of starting dates for those enrolling in a training program.

and that imposes the stochastic dominance assumption.

4.2.1 The Unrestricted Model

Table 4 presents estimated model parameters obtained by maximizing the likelihood function in (1). Each of the columns, labelled ML1, ML2, and ML3, corresponds to different local maximum likelihood (ML) points detected. The rows correspond to different parameters of the model or functions of them, such as the treatment effects of interest, and their corresponding standard errors. The estimated probability of being in group $G_i = g$ ($\hat{\pi}_g$), and the average potential outcome under treatment $Z_i = z$ for the individuals who participated in a language training program (the treated) and are in group $G_i = g$ ($\hat{AveTr}(g,z)$) are calculated, respectively, as:

$$\hat{\pi}_g = \sum_{i=1:N} \hat{\pi}_{g:i}/N$$

$$\hat{AveTr}(g,z) = \frac{\sum_{i\in(Z_i=1)} \hat{\pi}_{g:i} \exp(\mathbf{X}_i^{\mathrm{T}} \hat{\boldsymbol{\eta}}_{g,z} + 0.5 \cdot \sigma^2_{g,z})}{\sum_{i\in(Z_i=1)} \hat{\pi}_{g:i}}.$$

Several local solutions (modes) to the likelihood function in the unrestricted model were detected. As previously discussed, this is ascribed to the fact that the likelihood function of mixtures and models for truncated variables are non-regular. That is, their likelihood does not satisfy the regularity conditions for a likelihood to be symmetric, unimodal, and thus for the corresponding maximum likelihood (ML) estimator to be efficient (Lehmann and Casella, 1998, chapter 6). This implies that the likelihood can show multiple ML points even if it is identified (in the sense that the parameter space is in an one-to-one relation with the space of the model). Mercatanti (2013) shows that the likelihood for a closely related but simpler normal mixture model with non-compliance is identified but it only locally satisfies the regularity likelihood conditions and, consequently, it can exhibit multiple modes. Among the several local ML points detected in our model, Table 4 reports the extreme cases, corresponding to the lower and upper values of the estimated treatment effect on wages for the always-employed, and to the lower and upper values of the effect on employment. These cases are denoted in boldface.

The estimates in Table 4 indicate that the estimated proportion of individuals always employed is between 15% and 33% depending on the local ML point chosen. This is an important proportion since it reflects the size of the population for which we will estimate the effect of foreign language training programs on wages. Table 4 shows that the estimated effect on employment ($\hat{\pi}_{EN} - \hat{\pi}_{NE}$) is statistically significant but its sign, under these assumptions, depends on the local ML chosen: the lowest effect is estimated at –0.194 while the highest effect is estimated at +0.158. Similarly, the estimated effect of language training on the wages of those always employed is statistically significant but its sign depends on the local ML chosen: the lowest effect is estimated at –7.05 Euro per hour while the highest effect is estimated at +3.04 Euro per hour. Naturally, it is far from desirable that the sign of the treatment effects of interest depends on the local ML point chosen. Therefore, we proceed to include a secondary outcome in an attempt to regularize the likelihood function of the unrestricted model.

4.2.2 The Model with a Secondary Outcome

We employ the number of hours worked as a secondary outcome in order to improve inference on the parameters of interest in our model. To do this, we employ the likelihood function outlined in section 3.3. The rationale to include a secondary outcome follows the literature on the use of mixture models for causal inference (Mattei et al., 2013; Mealli and Pacini, 2013; Mercatanti et al., 2015), which shows that a secondary outcome can greatly improve the inference for the primary outcome by providing extra information to predict the mixture

Table 4: Some local ML estimates detected for the unrestricted model.

	ML1		ML2		ML3	
$\hat{\pi}_{EE}$	0.153	(.004)	0.174	(.004)	0.332	(.004)
$\hat{\pi}_{EN}$	0.225	(.011)	0.124	(.013)	0.312	(.010)
$\hat{\pi}_{NE}$	0.336	(.004)	0.318	(.004)	0.154	(.004)
$\hat{\pi}_{NN}$	0.286	(.011)	0.382	(.013)	0.201	(.010)
Est. effect on employment: $\hat{\pi}_{EN} - \hat{\pi}_{NE}$	-0.111	(.012)	**–0.194**	(.013)	**0.158**	(.011)
$Ave\hat{Tr}(EE, 1)$	14.77	(.17)	15.77	(.34)	14.95	(.19)
$Ave\hat{Tr}(EE, 0)$	21.82	(.32)	20.90	(.28)	11.91	(.04)
$Ave\hat{Tr}(EN, 1)$	16.18	(.52)	25.09	(.04)	13.09	(.68)
$Ave\hat{Tr}(NE, 0)$	11.92	(.04)	11.80	(.42)	21.71	(.32)
Est. effect on hourly wages for treated EE	**–7.05**	(.34)	-5.13	(.42)	**3.04**	(.19)
log-Likelihood	$-15,555$		$-15,630$		$-15,726$	

Boldface indicates the lower and upper values for the treatment effects on employment and wages, which was the basis for choosing the local ML points presented in the table. Standard errors are shown in parentheses.

membership and disentangle the mixtures.

Table 5 presents the estimates with a secondary outcome, which includes an additional ML point to continue presenting the lower and upper estimated values of the parameters of interest. Table 5 shows that the introduction of the secondary outcome improves inference by reducing the range of extreme values for both estimated effects of interest. This reduction is more notorious on the estimated effect on wages for the always-employed than on the effect on employment. The estimated effect of language training on employment is statistically significant and the lowest and highest effects across local ML points are estimated at –0.150 and +0.171, respectively. The estimated effect of language training on the wages of those always employed is also statistically significant and the lowest and highest effects across local ML points are estimated at –3.49 and +2.39, respectively. The range of estimates in Table 5 for the proportion of individuals always employed is considerably reduced by between 20 and 29 percent, depending on the local ML point chosen.

Table 5: Some local MLEs detected for the model with secondary outcome.

	ML1		ML2		ML3		ML4	
$\hat{\pi}_{EE}$	0.286	(.003)	0.209	(.003)	0.200	(.003)	0.280	(.003)
$\hat{\pi}_{EN}$	0.170	(.011)	0.133	(.013)	0.235	(.010)	0.377	(.009)
$\hat{\pi}_{NE}$	0.204	(.003)	0.283	(.003)	0.288	(.003)	0.206	(.003)
$\hat{\pi}_{NN}$	0.340	(.011)	0.374	(.013)	0.277	(.011)	0.136	(.009)
Est. effect on employment $\hat{\pi}_{EN} - \hat{\pi}_{NE}$	-0.034	(.012)	**–0.150**	(.014)	-0.053	(.011)	**0.171**	(.010)
$Ave\hat{Tr}(EE, 1)$	12.34	(.49)	13.07	(.52)	15.34	(.55)	14.51	(.54)
$Ave\hat{Tr}(EE, 0)$	15.83	(.11)	13.03	(.09)	12.95	(.09)	15.80	(.11)
$Ave\hat{Tr}(EN, 1)$	17.84	(1.07)	13.65	(.99)	15.11	(.54)	13.32	(.66)
$Ave\hat{Tr}(NE, 0)$	13.01	(.09)	15.88	(.11)	15.83	(.11)	12.94	(.09)
Est. effect on hourly wages for treated EE	**–3.49**	(.50)	0.04	(.53)	**2.39**	(.57)	-1.29	(.58)
log-Likelihood	$-79,088$		$-79,188$		$-79,159$		$-79,243$	

Boldface indicates the lower and upper values for the treatment effects on employment and wages, which was the basis for choosing the local ML points presented in the table. Standard errors are shown in parentheses.

4.2.3 The Model with a Secondary Outcome and the Stochastic Dominance Restriction

To further improve inference on the parameters of interest, we impose the stochastic dominance restriction in Assumption 3. This restriction can be reasonably advocated under the notion of a positive selection into employment. That is, factors that increase the individual's wage also increase her likelihood of working, which is implied by standard models of labor supply (e.g., Killingsworth, 1983). Following this intuition, the wage distribution for the always-employed (EE) when trained stochastically dominates that of the group of individuals that work only when trained (EN). Similarly, positive selection into employment implies that the wage distribution for the always-employed (EE) when not trained stochastically dominates that of the group of individuals that work only when not trained. Recent work placing bounds on the effects of different policies on wages has employed similar assumptions that are also justified based on positive selection into employment. Examples are Blundell et al. (2007), Lechner and Melly (2010), and Blanco et al. (2013).

Table 6 shows that the combined use of a secondary outcome and the stochastic dominance restriction improves inference considerably. The resulting likelihood function does not have a unique set of ML estimates, but it only exhibits three local ML points. Looking across the set of ML estimates in Table 6, the estimated proportion of individuals always employed is now very similar at 37%. The estimated proportion of individuals that are employed only if not trained (NE) is also very similar across ML points, while the estimated proportions of those never employed (NN) and those employed only if trained (EN) are a little more variable due to their estimates in one local point (ML2). All estimated proportions are highly statistically significant. The estimated effect of language training on employment is positive across the three local ML points. The first local ML point estimates this effect to be a statistically insignificant 0.008, while the other two ML points show highly significant estimates of 0.127 and 0.052, respectively. As for the estimated effect of language training on the wages of those always employed they are all negative, small, and statistically insignificant, ranging from –0.16 to –0.25 Euro per hour.

In sum, the combination of the use of a secondary outcome and the stochastic dominance

Table 6: The three local MLEs detected with secondary outcome and the stochastic dominance restriction.

	ML1	ML2	ML3
$\hat{\pi}_{EE}$	0.375 (.003)	0.375 (.003)	0.377 (.003)
$\hat{\pi}_{EN}$	0.123 (.009)	0.241 (.010)	0.166 (.010)
$\hat{\pi}_{NE}$	0.115 (.002)	0.114 (.002)	0.114 (.002)
$\hat{\pi}_{NN}$	0.386 (.009)	0.269 (.011)	0.342 (.010)
Est. effect on employment $\hat{\pi}_{EN} - \hat{\pi}_{NE}$	0.008 (.009)	0.127 (.011)	0.052 (.010)
$Ave\hat{Tr}(EE, 1)$	14.80 (.39)	14.73 (.39)	14.84 (.36)
$Ave\hat{Tr}(EE, 0)$	15.05 (.07)	14.99 (.07)	15.00 (.07)
$Ave\hat{Tr}(EN, 1)$	14.21 (.75)	11.91 (.82)	10.23 (10.11)
$Ave\hat{Tr}(NE, 0)$	13.47 (.13)	13.38 (.13)	13.38 (.13)
Est. effect on hourly wages for treated EE	-0.25 (.40)	-0.24 (.39)	-0.16 (.36)
log-Likelihood	$-82,698$	$-82,700$	$-82,731$

Standard errors are shown in parentheses.

assumption results in a much better behaved likelihood function for our model. The likelihood function still exhibits three local ML points, but their corresponding estimates across them do not change considerably. The models' results imply that language training programs likely have a positive and significant effect on the employment of participants, ranging from 5.2 to 12.7 percentage points. Considering the average employment rate of 51% from Table 1, the effect represents an increase in employment of between 10% and 25%. Conversely, language training programs appear to not have a significant effect on the wages of the individuals that are always employed (regardless of language training participation).

5 Sensitivity Analysis

In this section, we implement a sensitivity analysis with the goal of gauging the robustness of the main results in the previous section to violations of the key unconfoundedness assumption employed to identify the causal effects of interest. The general intuition behind the sensitivity analysis we employ is to assess the plausible impact of unmeasured confounders that lead to violations of the unconfoundedness assumption. This type of sensitivity analysis is rooted on similar analyses proposed in the context of other causal models in, for example, Rosenbaum and Rubin (1983b), Rosenbaum (2002), and Imbens (2003).

We concentrate on the model with secondary outcome and under the stochastic dominance assumption. Note that under unconfoundedness we have that $P(G_i|X_i, Z_i = 1) = P(G_i|X_i, Z_i = 0)$, where, as before, G_i denotes the principal strata. Thus, an implication of the failure of unconfoundedness is that $P(G_i|X_i, Z_i = 1) \neq P(G_i|X_i, Z_i = 0)$. We exploit this insight to assess the effects of interpretable unmeasured confounders by considering non-zero values of sensitivity parameters ξ_g ($g = EE, EN, NE$) for each of the strata (the NN stratum will be set, without loss of generality, as the base category below). These sensitivity parameters will alter the equality above that holds under unconfoundedness. Importantly, considering unmeasured confounders for the wage outcome is not necessary in this setting because their consequences cannot be distinguished from the effect of the treatment since we are not imposing an exclusion restriction assumption (Schwartz, Li and Reiter, 2012). Thus, our sensitivity analysis encompasses consequences of unmeasured confounders on both effects of interest.

For simplicity, in the following discussion we will consider probabilities that are not conditional on X_i. We consider values of the sensitivity parameter ξ_{EE} that decrease the probability to be always-employed in the treatment arm relative to control, that is, $P(G_i = EE|Z_i = 1) < P(G_i = EE|Z_i = 0)$. We interpret ξ_{EE} as an unobservable working toward the always-employed having less interest in the language training program, perhaps to free up time to look for a job. This is consistent with always-employed individuals having a strong preference for being employed. Meanwhile, we set ξ_{EN} to increase the probability to be EN in the treatment arm relative to control, that is, $P(G_i = EN|Z_i = 1) > P(G_i = EN|Z_i = 0)$. Presumably, EN are motivated to take the language training program (i.e., they may suspect they will remain unemployed otherwise), and thus ξ_{EN} may be interpreted as an unobservable (e.g., motivation) that increases the likelihood of enrolling in training. Lastly, we set ξ_{NE} to increase the probability to be NE in the treatment arm relative to control, that is, $P(G_i = NE|Z_i = 1) > P(G_i = NE|Z_i = 0)$. Intuitively, one can think of the NE as individuals that, when treated, would raise their reservation wage and reject employment that they would accept under control. Thus, ξ_{NE} may be interpreted as an unobservable (e.g., motivation) that increases the likelihood of enrolling in language training, as NE may strongly believe that training improves their skills (consistent with raising their future reservation wage).

To set plausible (and interpretable) values for the sensitivity parameters ξ_g, we concentrate on values for the difference in the conditional strata probabilities by treatment arm, denoted by $\Delta_g = P(G_i = g|Z_i = 1) - P(G_i = g|Z_i = 0)$. The values we consider are: $\Delta_{EE} \in \{0, -0.075, -0.15\}$, $\Delta_{EN} \in \{0, 0.05, 0.10\}$, and $\Delta_{NE} \in \{0, 0.05, 0.10\}$. This set of values results in $3^3 = 27$ different sensitivity scenarios for the set of six conditional strata probabilities by treatment arm $P(G_i|Z_i = z)$. They are obtained by adding or subtracting $(\Delta_g/2)$ to the corresponding marginal probabilities $P(G_i = g) = \pi_g$ previously estimated in the model with secondary outcome and under the stochastic dominance assumption, specifically those under ML3 shown in Table 6.[11] Note also that, for each difference, the value of zero is consistent

[11]For instance, denoting by $\pi_{g,ML3}$ the corresponding marginal probabilities, the conditional probabilities by treatment arm are obtained as $P(G_i = g|Z_i = 1) = \pi_{g,ML3} + (\Delta_g/2)$ and $P(G_i = g|Z_i = 0) = \pi_{g,ML3} - (\Delta_g/2)$.

with the validity of the unconfoundedness assumption for that stratum.

To gauge the relative importance of the departures from unconfoundedness considered in the sensitivity analysis, we refer to the estimated marginal probabilities π_g under the model with secondary outcome and under the stochastic dominance assumption for the local point ML3 in Table 6. For each Δ_g value, the following present the percentage, relative to the corresponding π_g in Table 6, that each value represents: $\Delta_{EE} \in \{0, -20\%, -40\%\}$, $\Delta_{EN} \in \{0, 30\%, 60\%\}$, and $\Delta_{NE} \in \{0, 44\%, 88\%\}$. Thus, our set of sensitivity scenarios represents up to a substantial departure from the equality in the conditional strata probabilities by treatment arm (the consequence of unmeasured confounders) relative to the estimated marginal probabilities under the validity of unconfoundedness.

The next step is to calculate, for each of the 27 scenarios under consideration, the implied sensitivity parameters ξ_g. To do this, we apply the log of odds-ratio formula: $\xi_g = log(P(G_i = g|Z_i = 1) \times P(NN|Z_i = 0)/P(G_i = g|Z_i = 0) \times P(G = NN|Z = 1))$ for $g = EE, EN, NE$. Therefore, each resulting value of the vector $(\xi_{EE}, \xi_{EN}, \xi_{NE})$ corresponds to one scenario $(\Delta_{EE}, \Delta_{EN}, \Delta_{NE})$. Once the implied sensitivity parameters ξ_g have been obtained in this way, they are included in the model for the principal strata membership:

$$P(G_i = g) = \pi_{g:i} = \frac{\exp(\mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta}_g + \xi_g)}{\sum_{g'} \exp(\mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta}_{g'} + \xi_{g'})},$$

and the corresponding likelihood function for the full model with secondary outcome and under the stochastic dominance assumption is maximized.[12] We note that, while our approach to sensitivity analysis is tractable and easily interpretable (in the context of our application), it represents an approximation in that we employ strata probabilities by treatment arm that are

[12]To decrease the computational burden in the estimation of the model parameters under the different sensitivity scenarios (and after checking that the same relation holds for a set of the sensitivity scenarios), the search of ML points is conducted by detecting the three local ML points under unconfoundedness and then conducting a search around them. No additional local ML points were detected throughout the sensitivity analysis.

not conditional on X. Accounting for this conditioning would substantially complicate the procedure.[13]

We summarize in what follows the results of the sensitivity analysis outlined above. Table 7 to Table 9 present the estimated model parameters from the sensitivity analysis for the local ML3 point under unconfoundedness in Table 6. Each of these tables shows, for a given value of Δ_{EE}, the estimated strata probabilities, average wages, and the effects on employability and wages (along with their standard errors), that correspond to each value of the couple $(\Delta_{EN}, \Delta_{NE})$. A complete set of tables showing the corresponding estimates under each one of the sensitivity scenarios for the other two ML points detected in Table 6 (ML1 and ML2) is available in the online appendix to the paper (these tables show similar results to those described here). In general, the values of $\hat{\pi}_{EE}$, $\hat{\pi}_{NE}$, $AveTr(\hat{E}E, 0)$ and $AveTr(\hat{N}E, 0)$ show only small differences across sensitivity scenarios. Therefore, the sensitivity observed in the estimated values of the effect on employability and wages, $(\hat{\pi}_{EN} - \hat{\pi}_{NE})$ and $Est.Eff.EE = AveTr(\hat{E}E, 1) - AveTr(\hat{E}E, 0)$, respectively, are due to the sensitivity observed in $\hat{\pi}_{EN}$ and $AveTr(\hat{E}E, 1)$. This may be due to the different entanglement of mixtures involved in the likelihood function. The estimated standard errors are fairly stable across sensitivity scenarios.

Table 10 and Table 11 summarize the degree of sensitivity of the causal effects of interest. They show, for each local ML point, the minimun and maximum estimated effect on employability (Table 10) and wages (Table 11), along with their standard errors (in parentheses) and the sensitivity scenario where they occur. Table 10 shows that the effect on employment is relatively robust to the departures from unconfoundedness reflected in the sensitivity scenarios, particularly for local points ML2 and ML3. For these two local points, the minimum

[13]For instance, in the calculation of the implied sensitivity parameters, instead of using the log of odds-ratio formula, accounting for the conditioning on covariates would involve a complicated differential equation.

Table 7: Sensitivity analysis estimates for ML3, $\Delta_{EE} = 0$

	$\hat{\pi}_{EE}$	$\hat{\pi}_{EN}$	$\hat{\pi}_{NE}$	$\hat{\pi}_{NN}$	$\hat{\pi}_{EN} - \hat{\pi}_{NE}$	$Ave\hat{Tr}(EE,1)$	$Ave\hat{Tr}(EE,0)$	Est.Eff.EE	$Ave\hat{Tr}(EN,1)$	$Ave\hat{Tr}(NE,0)$
$\Delta_{EN} = 0$	0.377	0.166	0.114	0.342	0.052	14.84	15.00	-0.16	10.23	13.38
$\Delta_{NE} = 0$	(.003)	(.010)	(.002)	(.010)	(.010)	(.36)	(.07)	(.36)	(1.11)	(.13)
$\Delta_{EN} = 0$	0.377	0.168	0.115	0.340	0.053	14.94	14.99	-0.05	10.28	13.38
$\Delta_{NE} = 0.05$	(.003)	(.010)	(.002)	(.010)	(.010)	(.36)	(.07)	(.36)	(1.11)	(.13)
$\Delta_{EN} = 0$	0.377	0.175	0.117	0.331	0.058	15.09	14.99	0.10	10.41	13.38
$\Delta_{NE} = 0.10$	(.003)	(.010)	(.002)	(.010)	(.010)	(.36)	(.07)	(.36)	(1.11)	(.13)
$\Delta_{EN} = 0.05$	0.377	0.153	0.114	0.355	0.039	14.89	15.00	-0.11	10.32	13.38
$\Delta_{NE} = 0$	(.003)	(.010)	(.002)	(.010)	(.010)	(.36)	(.07)	(.36)	(1.10)	(.13)
$\Delta_{EN} = 0.05$	0.377	0.155	0.115	0.352	0.040	14.99	15.00	-0.01	10.37	13.39
$\Delta_{NE} = 0.05$	(.003)	(.010)	(.002)	(.010)	(.010)	(.37)	(.07)	(.37)	(1.11)	(.13)
$\Delta_{EN} = 0.05$	0.377	0.162	0.117	0.344	0.045	15.14	14.99	0.15	10.50	13.39
$\Delta_{NE} = 0.10$	(.003)	(.010)	(.002)	(.010)	(.010)	(.37)	(.07)	(.38)	(1.13)	(.13)
$\Delta_{EN} = 0.10$	0.377	0.141	0.114	0.367	0.027	14.93	15.00	-0.07	10.40	13.39
$\Delta_{NE} = 0$	(.003)	(.009)	(.002)	(.009)	(.009)	(.36)	(.07)	(.37)	(1.08)	(.13)
$\Delta_{EN} = 0.10$	0.377	0.144	0.116	0.364	0.028	15.04	15.00	-0.06	10.47	13.39
$\Delta_{NE} = 0.05$	(.003)	(.009)	(.002)	(.009)	(.009)	(.37)	(.07)	(.37)	(1.09)	(.13)
$\Delta_{EN} = 0.10$	0.377	0.151	0.117	0.355	0.034	15.19	14.99	0.20	10.61	13.39
$\Delta_{NE} = 0.10$	(.003)	(.010)	(.002)	(.010)	(.010)	(.38)	(.07)	(.38)	(1.11)	(.13)

$\Delta_{EE} = P(G_i = EE|Z_i = 1) - P(G_i = EE|Z_i = 0)$

$\Delta_{EN} = P(G_i = EN|Z_i = 1) - P(G_i = EN|Z_i = 0)$

$\Delta_{NE} = P(G_i = NE|Z_i = 1) - P(G_i = NE|Z_i = 0)$

Table 8: Sensitivity analysis estimates for ML3, $\Delta_{EE} = -0.075$

	$\hat{\pi}_{EE}$	$\hat{\pi}_{EN}$	$\hat{\pi}_{NE}$	$\hat{\pi}_{NN}$	$\hat{\pi}_{EN} - \hat{\pi}_{NE}$	$Ave\hat{Tr}(EE,1)$	$Ave\hat{Tr}(EE,0)$	Est.Eff.EE	$Ave\hat{Tr}(EN,1)$	$Ave\hat{Tr}(NE,0)$
$\Delta_{EN} = 0$	0.376	0.186	0.114	0.323	0.072	14.89	15.00	-0.11	10.70	13.38
$\Delta_{NE} = 0$	(.003)	(.011)	(.002)	(.011)	(.011)	(.38)	(.07)	(.39)	(1.11)	(1.12)
$\Delta_{EN} = 0$	0.376	0.186	0.114	0.322	0.072	14.99	15.00	-0.11	10.79	13.38
$\Delta_{NE} = 0.05$	(.003)	(.011)	(.002)	(.011)	(.011)	(.38)	(.07)	(.39)	(1.11))	(1.12)
$\Delta_{EN} = 0$	0.376	0.193	0.116	0.314	0.077	15.14	14.99	0.15	10.86	13.38
$\Delta_{NE} = 0.10$	(.003)	(.011)	(.002)	(.011)	(.011)	(.38)	(.07)	(.39)	(1.10)	(1.12)
$\Delta_{EN} = 0.05$	0.376	0.169	0.114	0.341	0.055	14.94	15.00	-0.06	10.72	13.38
$\Delta_{NE} = 0$	(.003)	(.011)	(.002)	(.011)	(.011)	(.39)	(.07)	(.39)	(1.11)	(1.12)
$\Delta_{EN} = 0.05$	0.376	0.170	0.115	0.338	0.055	15.04	15.00	0.06	10.83	13.37
$\Delta_{NE} = 0.05$	(.003)	(.010)	(.002)	(.011)	(.011)	(.39)	(.07)	(.39)	(1.10)	(1.12)
$\Delta_{EN} = 0.05$	0.376	0.178	0.116	0.329	0.062	15.20	14.99	.21	10.91	13.37
$\Delta_{NE} = 0.10$	(.003)	(.010)	(.002)	(.010)	(.010)	(.39)	(.07)	(.39)	(1.09)	(1.12)
$\Delta_{EN} = 0.10$	0.376	0.153	0.114	0.356	0.039	14.98	15.00	-0.02	10.76	13.37
$\Delta_{NE} = 0$	(.003)	(.010)	(.002)	(.010)	(.010)	(.39)	(.07)	(.39)	(1.10)	(1.13)
$\Delta_{EN} = 0.10$	0.376	0.155	0.115	0.353	0.040	15.09	15.00	0.09	10.88	13.37
$\Delta_{NE} = 0.05$	(.003)	(.010)	(.002)	(.010)	(.010)	(.40)	(.07)	(.40)	(1.09)	(1.13)
$\Delta_{EN} = 0.10$	0.376	0.162	0.117	0.344	0.045	15.25	14.99	0.26	10.95	13.37
$\Delta_{NE} = 0.10$	(.003)	(.010)	(.002)	(.010)	(.010)	(.40)	(.07)	(.40)	(1.09)	(1.13)

$\Delta_{EE} = P(G_i = EE|Z_i = 1) - P(G_i = EE|Z_i = 0)$

$\Delta_{EN} = P(G_i = EN|Z_i = 1) - P(G_i = EN|Z_i = 0)$

$\Delta_{NE} = P(G_i = NE|Z_i = 1) - P(G_i = NE|Z_i = 0)$

Table 9: Sensitivity analysis estimates for ML3, $\Delta_{EE} = -0.15$

	$\hat{\pi}_{EE}$	$\hat{\pi}_{EN}$	$\hat{\pi}_{NE}$	$\hat{\pi}_{NN}$	$\hat{\pi}_{EN} - \hat{\pi}_{NE}$	$Ave\hat{Tr(EE,1)}$	$Ave\hat{Tr(EE,0)}$	Est.Eff.EE	$Ave\hat{Tr(EN,1)}$	$Ave\hat{Tr(NE,0)}$
$\Delta_{EN} = 0$	0.375	0.222	0.115	0.287	0.107	14.97	15.00	-0.03	11.13	13.36
$\Delta_{NE} = 0$	(.002)	(.010)	(.002)	(.010)	(.010)	(.38)	(.07)	(.39)	(.78)	(.13)
$\Delta_{EN} = 0$	0.376	0.217	0.115	0.292	0.102	15.05	15.00	0.05	11.31	13.36
$\Delta_{NE} = 0.05$	(.002)	(.010)	(.002)	(.010)	(.010)	(.38)	(.07)	(.39)	(.79)	(.13)
$\Delta_{EN} = 0$	0.376	0.223	0.116	0.285	0.107	15.20	15.00	0.20	11.42	13.36
$\Delta_{NE} = 0.10$	(.002)	(.010)	(.002)	(.010)	(.010)	(.38)	(.07)	(.39)	(.79)	(.13)
$\Delta_{EN} = 0.05$	0.376	0.196	0.114	0.314	0.082	15.00	15.00	0.00	11.28	13.36
$\Delta_{NE} = 0$	(.002)	(.010)	(.002)	(.010)	(.010)	(.38)	(.07)	(.39)	(.80)	(.13)
$\Delta_{EN} = 0.05$	0.376	0.195	0.115	0.314	0.080	15.11	15.00	0.11	11.36	13.36
$\Delta_{NE} = 0.05$	(.002)	(.010)	(.002)	(.010)	(.010)	(.38)	(.07)	(.39)	(.83)	(.13)
$\Delta_{EN} = 0.05$	0.376	0.203	0.116	0.305	0.087	15.28	15.00	0.28	11.42	13.36
$\Delta_{NE} = 0.10$	(.002)	(.010)	(.002)	(.010)	(.010)	(.38)	(.07)	(.39)	(.82)	(.13)
$\Delta_{EN} = 0.10$	0.376	0.173	0.114	0.337	0.059	15.07	15.01	0.06	11.20	13.36
$\Delta_{NE} = 0$	(.002)	(.010)	(.002)	(.010)	(.010)	(.38)	(.07)	(.38)	(.85)	(.13)
$\Delta_{EN} = 0.10$	0.376	0.199	0.115	0.310	0.084	14.84	14.99	-0.15	11.72	13.36
$\Delta_{NE} = 0.05$	(.002)	(.010)	(.002)	(.010)	(.010)	(.38)	(.07)	(.38)	(.85)	(.13)
$\Delta_{EN} = 0.10$	0.376	0.207	0.116	0.301	0.091	14.99	14.99	0.00	12.09	13.36
$\Delta_{NE} = 0.10$	(.002)	(.010)	(.002)	(.010)	(.010)	(.38)	(.07)	(.38)	(.87)	(.13)

$\Delta_{EE} = P(G_i = EE|Z_i = 1) - P(G_i = EE|Z_i = 0)$

$\Delta_{EN} = P(G_i = EN|Z_i = 1) - P(G_i = EN|Z_i = 0)$

$\Delta_{NE} = P(G_i = NE|Z_i = 1) - P(G_i = NE|Z_i = 0)$

Table 10: Sensitivity analysis, range of estimated effects on employability

	min		max	
	$min(\hat{\pi}_{EN} - \hat{\pi}_{NE})$	$(\Delta_{EE}, \Delta_{EN}, \Delta_{NE})$	$max(\hat{\pi}_{EN} - \hat{\pi}_{NE})$	$(\Delta_{EE}, \Delta_{EN}, \Delta_{NE})$
ML1	-0.029 (.009)	$(0, 0.10, 0)$	0.039 (.008)	$(-0.15, 0, 0.10)$
ML2	0.082 (.010)	$(0, 0.10, 0)$	0.153 (.010)	$(-0.15, 0, 0.10)$
ML3	0.027 (.009)	$(0, 0.10, 0)$	0.107 (.010)	$(-0.15, 0, 0.10)$ and $(-0.15, 0, 0)$

estimated effect on employment is a statistically significant 0.027, while the other minimum and maximum estimates are all positive and statistically significant. Conversely, while the maximum estimate for local point ML1 is positive and statistically significant, the minimum estimate is of the opposite sign (-0.029) and also statistically significant. Thus, there is a lack of robustness for this effect under ML1, which in Table 6 was statistically insignificant.

Table 11 shows that the effect on wages for the always-employed is robust to the departures from unconfoundedness reflected in the sensitivity scenarios for all local ML points, since none of the estimated minimum or maximum effects are statistically significant (and they remain close to zero). Thus, it appears that not even when plausible unobserved confounders are considered the estimated effects on wages for the always-employed become statistically significant. Overall, with the exception of the local point ML1 for the employment effect, the sensitivity analysis points to an acceptable robustness of the estimated effects shown in Table 6 to plausible departures from the crucial unconfoundedness assumption employed for identification.

Table 11: Sensitivity analysis, range of estimated effects on wages

	min		max	
	$min(\hat{\pi}_{EN} - \hat{\pi}_{NE})$	$(\Delta_{EE}, \Delta_{EN}, \Delta_{NE})$	$max(\hat{\pi}_{EN} - \hat{\pi}_{NE})$	$(\Delta_{EE}, \Delta_{EN}, \Delta_{NE})$
ML1	-0.37 (.40)	$(0, 0.10, 0)$	0.02 (.40)	$(-0.15, 0, 0.10)$
ML2	-0.27 (.39)	$(-0.075, 0, 0)$	0.14 (.40)	$(-0.075, 0.10, 0.10)$
ML3	-0.016 (.36)	$(0, 0, 0)$	0.26 (.40)	$(-0.075, 0.10, 0.10)$

6 Discussion and Conclusions

We evaluated the effect of language training programs for the unemployed in Luxembourg using administrative data that spans the period January 2007 to October 2011. Our outcomes of interest are the probability of re-employment and hourly wages, both measured 18 months after entering unemployment. To deal with selection into participation in language training programs, our main identifying assumption is that, conditional on observable characteristics, participation is not related to the outcomes of interest. Moreover, we employ a principal stratification framework to deal with selection into employment when considering the hourly wage outcome, for which we estimate the effect on the principal strata of individuals that are employed regardless of language training participation. Thus, our model suitably accounts for the selection into employment problem within an observational study in which the unemployed are not randomly assigned to training, but instead training participation depends on the observable characteristics of the unemployed.

For estimation, a normal mixture model is maximized using the EM algorithm within a "traditional" maximum likelihood approach (McLachlan and Peel, 2000). The unrestricted model presents several modes due to the non-regularity of the likelihood function, as it is typical in normal mixture models. We demonstrate that the combination of using a secondary outcome (hours worked) and the introduction of a stochastic dominance assumption substantially sharpens inference within our model by reducing the problem of multimodality and reducing the range of values of the estimates across the remaining local optima. This finding should be useful to researchers implementing models that lead to a mixture-of-normals likelihood function. Lastly, we conducted a sensitivity analysis that allows us to assess the robustness of our main results to the potential presence of unmeasured confounders that would render our identification assumption invalid.

Our results indicate that the language training for the unemployed in Luxembourg likely has a positive effect on the probability of employment, although in one set of results out of three (that correspond to local maximum likelihood points) we found no effect of the program on employment. Thus, the estimated effects of language training on employment range from no effect to an increase of up to 12.7 percentage points. As for the estimated effect of the language training program on the wages of those who would be employed regardless of training participation, we find no evidence of a statistically significant effect. The estimated effects on both outcomes are shown to be largely robust to a set of different values chosen for sensitivity parameters that model the potential presence of unmeasured confounders.

From a policy perspective, these findings suggest that the language training program likely has been successful in augmenting the re-employment probability of the unemployed. At the same time, it appears that language training programs do not to have noticeable effects on wages. There are at least three reasons that can explain this result. First, it may be that the language training programs in Luxembourg do not provide substantial human capital to the trainees and as a result their wages do not increase significantly. However, it is hard to argue that little human capital is formed under a 5-month average duration language training program that is certified by Luxembourg's national language institute. Second, it may be that the language training programs are made available to low-skilled and/or immigrant populations for whom wages are low in Luxembourg regardless of training. Indeed, there is evidence that Portuguese immigrants in Luxembourg are segregated at the level of economic sector of employment (Bulletin du Statec, 2012). Third, it could be that the training program does provide valuable human capital to participants, but that such human capital is more important for re-employment than for the level of compensation. Indeed, based on private conversations with ADEM officials, they seem to regard language training programs as instruments that remove language limitations to achieve re-employment. In this way, the human capital formed would not command a premium since it constitutes a necessary condition for employment in the majority of jobs. Still, it is of interest to examine in more detail the factors that may be behind the findings documented herein, particularly the relative importance of the second and third potential explanations just described.

References

[1] Aitkin, M. and Rubin, D. (1985). Estimation and hypothesis testing in finite mixture models. *Journal of the Royal Stcatistal Society - Series B*, 47, 67-75.

[2] Angrist, J., Imbens, G., and Rubin, D. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, 91, 444-455.

[3] Angrist, J., and Krueger., A. (1999). Empirical Strategies in Labor Economics, in Handbook of Labor Economics, vol. 3A, O. Ashen-felter and D. Card, ed.s. Amsterdam: North-Holland.

[4] Blanco, G., Flores-Lagunes, A., Flores C.A. (2013). Bounds on the Quantile Treatment Effects of Job Corps on Wages. *Journal of Human Resources*, 483, 659-701.

[5] Blundell, R., Gosling, A., Ichimura, H., and Meguir, C. (2007). Changes in the Distribution of Male and Female Wages Accounting for Employment Composition Using Bounds. *Econometrica*, 75(2): 387-399.

[6] Brosius, J. and Zanardelli, M. (2012). Evaluation de l efficacité des mesures de formation destinées aux chômeurs. LISER, Technical Report, Ministry of Labour, Employment, Social Economy and Economic Solidarity of Luxembourg.

[7] Bulletin du Statec, (2012). Statec, Luxembourg.

[8] Church, J. and King, I. (1993). Bilingualism and Network Externalities. *Canadian Journal of Economics*, Canadian Economics Association, 26, 337-45.

[9] Ciuperca, G., Ridolfi, A., and Idier, J. (2003). Penalized maximum likelihood estimator for normal mixtures. *Scandinavian Journal of Statistics*, 30, 45-59.

[10] Clausen, J., Hummelgaard, H., Husted, L., Blume, H. and Rosholm, M. (2006). The Impact of the Introduction Programme on the Labour-Market Integration. Copenhagen: Institute for Local Studies.

[11] Clausen, J., Hummelgaard, H., Husted, L., Blume K.J., and Rosholm M. (2006). The Impact of the Introduction Programme on the Labour-Market Integration. Copenhagen: Institute for Local Studies.

[12] Dempster, A., Laird, N., and Rubin, D. (1977) Maximum Likelihood from incomplete data via the EM algorithm (with discussion) *Journal of the Royal Statistical Society - Series B*, 39, 1-38.

[13] Donado, A. (2017). Foreign Languages and their Impact on Unemployment. *Labour*, 31(3), 265-287.

[14] Dustmann, C. and Fabbri, F. (2003). Language Proficiency and Labour Market Performance of Immigrants in the UK. *The Economic Journal*, 113, 695-717.

[15] European Social Fund in Luxembourg (2015). Report of the European Commission.

[16] Frangakis, C. and Rubin, D. (2002). Principal stratification in causal inference, *Biometrics*, 58, 21-29.

[17] Frumento, P., Mealli, F., Pacini, B. and Rubin, D. (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.

[18] Gerfin, M. and Lechner, M. (2002). A Microeconometric Evaluation of the Active Labour Market Policy in Switzerland. *The Economic Journal*, 112 (October), 854-893.

[19] Ginsburgh, V. and Juan Prieto-Rodriguez J. (2011). Returns to Foreign Languages of Native Workers in the EU. *Industrial and Labor Relations Review*, Cornell University, ILR School, 64, 599-618, April.

[20] Hathaway, R.J. (1985). A constrained formulation of maximum-likelihood estimation for normal mixture distributions. *The Annals of Statistics*, 13, 795-800.

[21] Heckman, J. (1979). Sample Selection Bias as a Specification Error. *Econometrica*, 47, 153-161.

[22] Heckman, J., LaLonde, R., and Smith, J. (1999). The Economics and Econometrics of Active Labor Market Programs, in Handbook of Labor Economics Vol.III, O. Ashenfelter, and D. Card, ed.s, pp. 18652097. Elsevier, Amsterdam.

[23] Imbens, G. (2003). Sensitivity to exogeneity assumptions in program evaluation. *The American Economic Review*, 93(2), 126-132.

[24] Imbens, W., and Angrist, J. (1994). Identification and Estimation of Local Average Treatment Effects. *Econometrica*, 62, 467-475.

[25] Isphording, I. (2013). Returns to Foreign Language Skills of Immigrants in Spain. *Labour*, CEIS, 27, 443-461.

[26] Isphording, I. (2014). Language and Labor Market Success. IZA working paper DP No. 8572.

[27] Isphording, I. and Otten, S. (2013). The Costs of BabylonLinguistic Distance in Applied Economics. *Review of International Economics*, Wiley Blackwell, 21, 354-369.

[28] Killingsworth, M. (1983). Labor Supply. Cambridge Surveys of Economic Literature. Cambridge University Press.

[29] Kuhn, P. and Lozano, F. (2008). The Expanding Workweek? Understanding Trends in Long Work Hours among U.S. Men, 1979-2006. *Journal of Labor Economics*, 26(2), 311-343.

[30] Lechner, M. (1999). Earnings and Employment Effects of Continuous Off-the-Job Training in East Germany After Unification. *Journal of Business and Economic Statistics*, 17, 74-90.

[31] Lechner, M. and Melly, B. (2010). Partial Identification of Wage Effects of Training Programs. University of St. Gallen. Unpublished.

[32] Lechner, M., and Wunsch, C. (2013). Sensitivity of Matching-Based Program Evaluations to the Availability of Control Variables. *Labour Economics*, 21, 111-121.

[33] Lee, D. (2009). Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects. *Review of Economic Studies*, 76, 10711102.

[34] Lehmann, E., and Casella, G. (1998). Theory of point estimation, 2nd edn. New York: Springer-Verlag.

[35] Lohmann, J. (2011). Do language barriers affect trade? *Economics Letters*, Elsevier, 110, 159-162.

[36] 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*, 4, 2336-2360.

[37] McHugh M. and Challinor, A. (2011). Improving Immigrants Employment Prospects through Work-Focused Language Instruction. Migration Policy Institute (US).

[38] McLachlan, G., and Peel, D. (2000). Finite Mixture Model. New York: Wiley.

[39] Mealli F., and Pacini B., (2013). Using secondary outcomes to sharpen inference in randomized experiments with noncompliance. *Journal of the American Statistical Association*, 108, 1120-1131.

[40] Mercatanti, A. (2013). A likelihood-based analysis for relaxing the exclusion restriction in randomized experiments with nonmcompliance. *Australian and New Zealand Journal of Statistics*, 55, 129-153.

[41] Mercatanti, A., Li, F., and Mealli, F. (2015). Improving inference of Gaussian mixtures using auxiliary variables. *Statistical Analysis and Data Mining*, 8, 34-48.

[42] OECD (2007). Jobs for Immigrants , Volume 1: Labour Market Integration in Australia, Denmark, Germany and Sweden.

[43] OECD Economic Surveys: (2010, 2012).

[44] OECD Economic Surveys: Luxembourg (2012).

[45] OECD Jobs for Immigrants , Volume 1: Labour Market Integration in Australia, Denmark, Germany and Sweden. Paris: OECD.

[46] Rosenbaum, P. (2002). Observational studies. Second Edition. Springer-Verlag: New York.

[47] Rosenbaum, P. and Rubin, D. (1983a). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70, 41-55.

[48] Rosenbaum, P. and Rubin, D. (1983b) Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. *Journal of the Royal Statistical Society - -Series B*, 45, 212218.

[49] Rubin, D. (1974). Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies. *Journal of Educational Psychology*, 66, 688701.

[50] Rubin, D. (1976). Inference and missing data. *Biometrika*, 63, 581-592.

[51] Rubin, D. (1978). Bayesian inference for causal effects. *Annals of Statistics*, 6, 34-58.

[52] Rubin, D. (1980). Discussion of Randomization analysis of experimental data: the Fisher randomization test by D. Basu. *Journal of the American Statistical Association*, 75, 591-593.

[53] Selten, R. and Pool, J. (1991). The distribution of foreign language skills as a game equilibrium Game Equilibrium Models, 4, *Social and Political Interaction*, Springer.

[54] Statec (2012). Population by nationality.
http://www.statistiques.public.lu/en/news/population/population/2012/08/20120821/index.html

[55] Statec (2013). Luxembourg Un demi-siècle de constantes et de variables, ISBN : 978-2-87988-112-6, STATEC.

[56] Statec (2014a). Luxembourg en chiffres 2014 [Luxembourg in figures]. *http://www.statistiques.public.lu/catalogue-publications/luxembourg-enchiffres/luxembourg-chiffres.pdf*

[57] Statec (2014b). Note de conjoncture 2014-2 [Monthly report 2014-2]. *http://www.statistiques.public.lu/catalogue-publications/noteconjoncture/2014/PDF-NDC-02-14.pdf*

[58] Schwartz, S., Li, F., and Reiter, J. (2012). Sensitivity analysis for unmeasured confounding in principal stratification. *Statistics in Medicine*, 106, 1331-1344.

[59] Williams, D. (2011). Multiple language usage and earnings in Western Europe. *International Journal of Manpower*, Emerald Group Publishing, 32, 372-393.

[60] Zhang, J., Rubin, D., and Mealli, F. (2008). Evaluating teh Effects of Job Training Programs on Wages Through Principal Stratification, in D. Millimet, J. Smith and E. Vytlacil (eds.) Modelling and Evaluating Treatment Effects in Econometrics, Vol. 21 (Amsterdam: Elsevier) 119147.

[61] 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.