Evaluation of Language Training Programs in Luxembourg using Principal Stratification

Michela Bia, Alfonso Flores-Lagunes, Andrea Mercatanti

Observational Studies, Volume 8, Issue 1, 2022, pp. 1-44 (Article)

Published by University of Pennsylvania Press
DOI: *https://doi.org/10.1353/obs.2022.0001*

Observational Studies Vol. 8.1 (2022) Submitted Feb. 2022; Published May 2022

Evaluation of Language Training Programs in Luxembourg using Principal Stratification

Michela Bia **michela.bia@liser.lu**
Evaluation Unit
LISER Luxembourg Institute of Socio-Economic Research
Luxembourg

Alfonso Flores-Lagunes **afloresl@syr.edu**
Department of Economics and Center for Policy Research
Syracuse University, USA,
and
IZA, Bonn,
and
Global Labor Organization (GLO)

Andrea Mercatanti **andrea.mercatanti@univr.it**
Department of Economics,
University of Verona, Italy,
and
Global Labor Organization (GLO)

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 simultaneously, using high-quality administrative data from Luxembourg. We address selection into training with an unconfoundedness assumption 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 a sensitivity analysis to assess the unconfoundedness assumption. Our results 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.

Keywords: Language Training Programs, Policy Evaluation, Principal Stratification, Mixture Models, Unconfoundedness, Sensitivity Analysis.

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. In 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 (Lohmann, 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 (Dustmann and Fabbri, 2003; Williams, 2011; Ginsburgh and Prieto-Rodriguez, 2011; Isphording, 2013; Donado, 2017) that are also present throughout the earnings distribution and across occupations (Ginsburgh and Prieto-Rodriguez, 2011; Isphording, 2014).

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 (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 in the U.S. (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 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.

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 Commission, 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 would be 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 (from January 2007 to October 2011) that registered with ADEM and 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 similarly 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 a plausible assumption in light 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]

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 et al., 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.

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 (Heckman, 1979; Angrist and Krueger, 1999). However, finding variables that are related to employment but not related to hourly wages is 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).

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. The resulting likelihood function shows multimodality—a high number of local maxima that makes inference challenging—due to the presence of mixture of distributions in our model. However, we demonstrate that the number of local maxima can be 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 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. Our approach differs from the approach in those papers, in which the likelihood analyses are aimed at detecting the global maximum likelihood point (usually labeled maximum likelihood estimator or MLE) such as in frequentist likelihood analyses for regular models,[3] or in direct likelihood analyses (Seaman et al., 2013). We show that our proposed approach can considerably reduce the multimodality problem in a tractable way that allows analyzing and interpreting the few local maximum likelihood points detected. In this way, we avoid imposing stronger model restrictions on the variances or on the proportions of principal strata (Hathaway, 1985; Aitkin and Rubin, 1985) to achieve point identification. A final important practical aspect we consider in our observational setting is the implementation of a sensitivity analysis in the spirit of Rosenbaum (2012) 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, 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 joint adoption of a secondary outcome and a stochastic dominance restriction helps to overcome situations in which the regularity of the likelihood function is broken. Second, this study 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. Third, we contribute to the growing literature on the labour market effects of language skills (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. Fourth, 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 (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

3. That is, for models whose likelihood functions satisfy the Cramér (1946) and Wald (1949) conditions that ensure efficiency of the global maximum likelihood estimator. Examples are cases in which the data are supposed to be drawn from a distribution in the exponential family (Lehmann and Casella, 1998).

the effectiveness of an important component of the bundle of training programs offered by ADEM: language training.

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,b). 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 have become the largest foreign community in the country (Statec, 2012a). 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. 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.[4] 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 makers have to design interven-

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

tions 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. We focus here on unemployed individuals who exclusively enrolled in language training programs.[5]

ADEM implements a "personalized assistance" model tailored to the needs of the unemployed individual. 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, mobility status, job expertise, job prospects 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 to ADEM's case-workers is summarized in a pre-training 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 (Brosius and Zanardelli, 2012; OECD, 2010, 2012). At the same time, OECD (2010), OECD (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 available empirical evidence on the effectiveness of ALMPs in Luxembourg by focusing on language training programs. 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 1 if the unemployed individual receives language training, and 0 if he does not receive language training. Let $Y_{1,i}(z)$ and $Y_{2,i}(z)$ denote two sets of potential outcomes for individual i, namely the potential values of each of two outcomes under each of the two possible treatment assignments: $z = 0, 1$. $Y_{1,i}$ represents the hourly wage 18 months after entering ADEM, one of

5. The language courses by ADEM are certified by the "Institut National des Langues" (http://www.inll.lu/en/). 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 (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).

the two outcomes of primary interest. $Y_{2,i}$ represents the number of hours worked 18 months after entering ADEM, a variable that will be used below as a secondary outcome to improve inference. Let $\mathbf{X}_i$ be a k-vector of pre-treatment characteristics and S_i be a binary post-treatment variable, equal to 1 if subject i is employed 18 months after registration at ADEM, 0 otherwise. S_i is the second outcome of interest that determines the observability of $Y_{1,i}$ and that is likely affected by Z_i. We denote the potential values of this variable as a function of the treatment as $S_i(z)$. Let $\mathbf{Z}$, $\mathbf{S}$, $\mathbf{Y_1}$, and $\mathbf{Y_2}$ be N-dimensional vectors with ith elements equal to Z_i, S_i, $Y_{1,i}$ and $Y_{2,i}$ respectively; and let $\mathbf{X}$ be the $N \times k$ matrix of pretreatment variables with ith row equal to $\mathbf{X}_i$. The goal is to identify and estimate the causal effect of Z_i on both outcomes of interest, S_i and $Y_{1,i}$.

To identify the effect of Z_i on $Y_{1,i}$, 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. This issue relates the two outcomes of interest ($Y_{1,i}$ 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), a concept that will be formalized 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 $Y_{1,i}$ 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 for whom $S_i(z) = 1$ (the 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 estimand 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 $\pi_{EN} - \pi_{NE}$. 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: $E[Y_{1,i}(1) - Y_{1,i}(0)|EE]$. This is a commonly estimated parameter in the literature (Zhang et al., 2009; Lechner and Wunsch, 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): $Z_i \perp \{Y_{1,i}(0), Y_{1,i}(1), Y_{2,i}(0), Y_{2,i}(1), S_i(0), S_i(1)\}|\mathbf{X}_i$.

Assumption 2 (Overlap condition): $0 < \Pr(Z_i = 1|\mathbf{X}_i) < 1$, for all i.

Assumption 1 states that, conditional on observable variables $\mathbf{X}_i$, the treatment is independent of all pairs of potential outcomes. This assumption, although widely used in the literature (Heckman et al., 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 $\mathbf{X}_i$). Nevertheless, we deem this assumption tenable given the combination of access to rich administrative data and the institutional features of the assignment of unemployed individuals into language training programs. 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 language training program (conditional on $\mathbf{X}_i$) is bounded away from zero or one.

7. This estimand 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.

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

Assumption 3 (Stochastic Dominance): *For any real number t,* $P(Y_{r,EE}(1) \leq t) \leq P(Y_{r,EN}(1) \leq t)$ *and* $P(Y_{r,EE}(0) \leq t) \leq P(Y_{r,NE}(0) \leq t)$ *, where* $r = 1, 2$.

This assumption states that the distribution of the EE when trained stochastically dominates the distribution of the EN when trained, and that the distribution of the EE when not trained stochastically dominates the distribution of the NE when not trained. This condition is imposed on both, the wage distribution (Y_1), and the hours worked distribution (Y_2). 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 and hours worked 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. An important difference with Zhang et al. (2009), however, is that the covariates in $\mathbf{X}_i$ 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 models for the potential outcomes and the principal strata membership. The two causal effects of interest are simultaneously estimated along with other parameters. This approach requires the prediction of the individuals' missing membership to the principal strata. The membership is unknown since one potential value of $S_i(z)$ 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 $Y_{1,i}^{\text{mis}} = Y_{1,i}(z) : z \neq z_i^{\text{obs}}$. Because we condition the analysis on the empirical distribution of the pre-treatment variables, $Pr(\mathbf{X}_i)$ does not need to be modelled. Additionally, Assumptions 1 and 2 imply that we can ignore the assignment mechanism $Pr(Z_i|\mathbf{X}_i)$. Thus, we focus on the distribution of the potential quantities $Y_{1,i}(z)$ and $S_i(z)$ given the pre-treatment variables which, by integration over the missing quantities, yields the following likelihood:

$$\mathcal{L}\left(\boldsymbol{\theta}^S, \boldsymbol{\theta}^{Y_1}; \mathbf{Z}^{\text{obs}}, \mathbf{S}^{\text{obs}}, \mathbf{Y}_1^{\text{obs}}, \mathbf{X}\right) = \prod_i \left[\int\int Pr\Big(S_i(0), S_i(1)|\mathbf{X}_i; \boldsymbol{\theta}^S) \cdot \right.$$
$$\left. \cdot Pr\Big(Y_{1,i}(0), Y_{1,i}(1)|S_i(0), S_i(1), \mathbf{X}_i; \boldsymbol{\theta}^{Y_1}) dY_{1,i}^{\text{mis}} dS_i^{\text{mis}}\right]$$

where $\boldsymbol{\theta}^S$ and $\boldsymbol{\theta}^{Y_1}$ 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_1 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 models for wages (Y_1) are specified as log-normal and allowed to vary by treatment status (the subindex 0 or 1 on the parameters):

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

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:

$$
\begin{aligned}
&\mathcal{L}(\boldsymbol{\theta}|\mathbf{Z}, \mathbf{S}^{\text{obs}}, \mathbf{Y}_1^{\text{obs}}, \mathbf{X}) \propto \\
&\quad \prod_{i\in(Z_i=1,S_i^{\text{obs}}=1)} [\pi_{EE:i} P(\log[Y_{1,i}]|\mathbf{X}_i, G_i = EE, Z_i = 1) + \pi_{EN:i} P(\log[Y_{1,i}]|\mathbf{X}_i, G_i = EN, Z_i = 1] \times \\
&\quad \prod_{i\in(Z_i=1,S_i^{\text{obs}}=0)} (\pi_{NE:i} + \pi_{NN:i}) \times \\
&\quad \prod_{i\in(Z_i=0,S_i^{\text{obs}}=1)} [\pi_{EE:i} P(\log[Y_{1,i}]|\mathbf{X}_i, G_i = EE, Z_i = 0) + \pi_{NE:i} P(\log[Y_{1,i}]|\mathbf{X}_i, G_i = NE, Z_i = 0)] \times \\
&\quad \prod_{i\in(Z_i=0,S_i^{\text{obs}}=0)} (\pi_{EN:i} + \pi_{NN:i}) = \\
&\quad \prod_{i\in(Z_i=1,S_i^{\text{obs}}=1)} [\pi_{EE:i} N(\mathbf{X}_i^{\text{T}} \boldsymbol{\eta}_{EE,1}, \sigma^2_{EE,1}) + \pi_{EN:i} N(\mathbf{X}_i^{\text{T}} \boldsymbol{\eta}_{EN,1}, \sigma^2_{EN,1})] \times \\
&\quad \prod_{i\in(Z_i=1,S_i^{\text{obs}}=0)} (\pi_{NE:i} + \pi_{NN:i}) \times \\
&\quad \prod_{i\in(Z_i=0,S_i^{\text{obs}}=1)} [\pi_{EE:i} N(\mathbf{X}_i^{\text{T}} \boldsymbol{\eta}_{EE,0}, \sigma^2_{EE,0}) + \pi_{NE:i} N(\mathbf{X}_i^{\text{T}} \boldsymbol{\eta}_{NE,0}, \sigma^2_{NE,0})] \times \\
&\quad \prod_{i\in(Z_i=0,S_i^{\text{obs}}=0)} (\pi_{EN:i} + \pi_{NN:i})
\end{aligned} \tag{1}
$$

The maximization of (1) 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 both the principal strata and potential outcomes models parameters. Then, in the maximization step, the likelihood function, conditional on the expected stratum membership, 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 and estimands are obtained by relying on their asymptotic distribution using the outer product of gradients and the Delta method, respectively (McLachlan and Peel, 2000).[9] 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 (Seaman et al., 2013), 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 chal-

9. (McLachlan and Peel, 2000) recommend the use of the bootstrap since the sample size has to be very large for the asymptotic approximation to work well. We show some evidence in Section 4 suggesting that in our application the asymptotic approximation yields standard errors that are not very different from the bootstrap. Thus, we prefer the asymptotic approximation due to the computational intensity of the bootstrap in this context.

lenging. This feature of the likelihood function is caused by the high degree of the likelihood equations, as shown by Buot and Richards (2006) and Catanese et al. (2006) via the theory of polynomial equations. Moreover, Wald's (1949) conditions, which would guarantee the efficiency of the global maximum likelihood, are difficult to check for models involving multiple roots (Small et al., 2000). Consequently, given 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), testing the consistency of the detected roots (Gan and Jiang, 1999), imposing suitable constraints on the variances (Hathaway, 1985) or on the proportions of mixture components (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 these proposals in our model would considerably increase the computational burden, they would require analytically proving the generalization of the corresponding theoretical results to our complex model, and they would imply introducing potentially strong assumptions on the variances or on the probabilities of principal strata. For these reasons, we propose a different approach.

To ameliorate the multimodality problem, we employ a secondary outcome ($Y_{2,i}$) and also impose the restriction of stochastic dominance (Assumption 3). The latter enters the likelihood by imposing restrictions on the parameter vectors $\boldsymbol{\eta}$ and the variances σ^2 (Zhang et al., 2009). For the mixture concerning the units $i \in (Z_i = 1, S_i^{\text{obs}} = 1)$ in (1), the coefficients other than the intercept in $\boldsymbol{\eta}_{EE,1}$ are imposed to be equal to those in $\boldsymbol{\eta}_{EN,1}$, and $\sigma^2_{EE,1}$ is imposed to be equal to $\sigma^2_{EN,1}$. The intercept in $\boldsymbol{\eta}_{EE,1}$ is instead constrained to be no less than that in $\boldsymbol{\eta}_{EN,1}$. Similar arguments apply to the mixture concerning the units $i \in (Z_i = 0, S_i^{\text{obs}} = 1)$. The resulting likelihood function improves inference and the multimodality problem is considerably reduced to the point that the few local maximum likelihood points remaining can be interpreted and contextualized.

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. 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).[10] The labor economics literature documents a strong correlation between hourly wages and the number of hours worked (Kuhn and Lozano, 2008). Including the secondary outcome to improve inference, the potential outcomes model (in more compact notation) becomes:

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

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

where

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

The expanded outcome model is then inserted into the general formulation of the likelihood 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 wages, we combine two rich administrative datasets. The richness of the data, in particular the availability of relevant pre-treatment individual characteristics, is instrumental in arguing the plausibility of our identifying assumptions.

The first dataset contains 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 dataset 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. Therefore, we make inference on the population of unemployed individuals that register with ADEM (and can be linked to IGSS) by sampling on the time period just mentioned. This is a policy-relevant population as the country's ALMPs target precisely this group.

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. Information is also available 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 pre-training score variable assessing the employability level of the unemployed worker, which case-workers use 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 (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 53% of the unemployed who participated in a language training program are employed, relative to only 51% of those who did not participate in any training program. This unadjusted difference likely lacks causal interpretation since participation in language training programs is not randomly determined.

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

Tables A.1 and A.2 (in Appendix A) present summary statistics for selected variables in our sample. Table A.1 shows that about 53% of our sample consists of males, and about 49% of individuals are married. In the sample, 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. 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. The first two rows of Table A.2 describe the labour market attachment of individuals. For instance, they have worked, on average, 3.90 and 6.85 months out of the last 6 and 12 months, respectively. Finally, looking at the employability level—the pre-training 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.[11] Lastly, the average number of months prior to taking training is 3.38.[12]

11. This variable's category of "to be determined" reflects unemployed individuals for whom the ADEM's case-worker chose to delay assigning a value. Typically, this assignment is done at a later meeting of the individual with ADEM (still before treatment assignment), but unfortunately that determination is not available to us.
12. 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.

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 (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 and that imposes the stochastic dominance assumption.

4.2.1 THE UNRESTRICTED MODEL

Table 2 presents estimated model parameters obtained by maximizing (1). 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 estimands 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 $G_i = g$ are calculated, respectively, as:

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

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

Several local solutions (modes) to the likelihood function in the unrestricted model were detected. As previously discussed, this is ascribed to the high degree of the likelihood equations for mixture models (Buot and Richards, 2006; Catanese et al., 2006). 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 and, even if it satisfies Cramér (1946) regularity conditions, it can exhibit multiple modes. Among the several local ML points detected in our model, Table 2 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 2 indicate that the estimated proportion of individuals always employed is between 15% and 33% depending on the local ML point chosen. This proportion reflects the size of the population for which we will estimate the effect of foreign language training programs on wages. Table 2 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 estimated 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.

Table 2: 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)
$\hat{AT}(EE, 1)$	14.77	(.17)	15.77	(.34)	14.95	(.19)
$\hat{AT}(EE, 0)$	21.82	(.32)	20.90	(.28)	11.91	(.04)
$\hat{AT}(EN, 1)$	16.18	(.52)	25.09	(.04)	13.09	(.68)
$\hat{AT}(NE, 0)$	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.

4.2.2 THE MODEL WITH A SECONDARY OUTCOME

We employ the number of hours worked as a secondary outcome to improve inference as outlined in section 3.3. The rationale to include a secondary outcome follows the literature (Mattei et al., 2013; Mealli and Pacini, 2013; Mercatanti et al., 2015) showing that 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.

Table 3 presents the estimates with a secondary outcome, which requires an additional ML point to continue presenting the lower and upper estimated values of the parameters of interest. The introduction of the secondary outcome improves inference by reducing the range of extreme values for both estimands of interest. The range of estimates for the proportion of individuals always employed is considerably reduced to 20-29 percent, depending on the local ML point chosen. 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.

Table 3: 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)
$\hat{AT}(EE, 1)$	12.34 (.49)	13.07 (.52)	15.34 (.55)	14.51 (.54)
$\hat{AT}(EE, 0)$	15.83 (.11)	13.03 (.09)	12.95 (.09)	15.80 (.11)
$\hat{AT}(EN, 1)$	17.84 (1.07)	13.65 (.99)	15.11 (.54)	13.32 (.66)
$\hat{AT}(NE, 0)$	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, 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 and worked hours also increase her likelihood of working, which is implied by standard models of labor supply in economics (Killingsworth, 1983). The intuition is that the traits (observed and unobserved) that make an individual have higher earnings potential in the labor market (e.g., marketable skills, discipline, conscientiousness) will also increase her worked hours and likelihood of participating in the labor market. Following this intuition, under the receipt of training, the wage and worked hours distributions for the always-employed (EE) stochastically dominates that of the group of individuals that work only when trained (EN). Similarly, positive selection into employment implies that, under no training, the wage and worked hours distributions for the always-employed (EE) 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 stochastic dominance assumptions that are also justified based on positive selection into employment. (Blundell et al., 2007; Lechner and Melly, 2010; Blanco et al., 2013).

One can think of situations in which Assumption 3 may not hold, which are important since our inference can be biased if the assumption is violated.[13] One involves individuals employed only under training (EN) if they get access to jobs where language skills are required while always-employed (EE) individuals land lower-paying jobs where no language skills are needed. We note that EE individuals also receive language training under the assumption, and thus they would also

13. We thank an anonymous reviewer for these suggested situations.

qualify for jobs where language training skills are needed. Thus, we argue that given positive selection into employment, if EE individuals do not take these jobs is because they take jobs that are more highly paid (holding other factors constant like preference and luck). A second situation considers that individuals employed only under no language training (NE) could become engaged in a better program which results in higher pay. We argue that the institutional framework largely prevents this, since, given the nature of our data, we know that a NE individual does not undergo any other ADEM training program. And enrolling in an alternative training program is unlikely since it implies paying for it out-of-pocket while having registered with ADEM to gain access to unemployment benefits.

Table 4 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, 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.[14]

In sum, the combination of the use of a secondary outcome and the stochastic dominance 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 statistically 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 statistically significant effect on the wages of the individuals that are always employed regardless of language training participation.

14. As mentioned in Section 3.3, for these set of results, we also estimated standard errors using the bootstrap to compare them to the asymptotic approximation using the Delta method. Implementing the bootstrap in our context is very computationally intensive. For each of the detected ML points we run a bootstrap with 100 replications where we take the ML parameter vector calculated from the original data as the starting value in applying the EM algorithm to each bootstrap sample. We show in the Appendix Table B.3 that the standard errors computed using the Delta method and the bootstrap are fairly close to each other, and that the statistical conclusions achieved by the implied p-values are the same for all estimands. We take this as the basis for our decision to report standard errors based on the Delta method throughout the paper.

Table 4: 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)
$\hat{AT}(EE, 1)$	14.80	(.39)	14.73	(.39)	14.84	(.36)
$\hat{AT}(EE, 0)$	15.05	(.07)	14.99	(.07)	15.00	(.07)
$\hat{AT}(EN, 1)$	14.21	(.75)	11.91	(.82)	10.23	(10.11)
$\hat{AT}(NE, 0)$	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.

5. Sensitivity Analysis

In this section, we implement a sensitivity analysis with the goal of gauging the robustness of main results to violations of the key unconfoundedness assumption employed for identification. The general intuition is to assess the plausible impact of unmeasured confounders that lead to violations of the unconfoundedness assumption. This type of sensitivity analysis is rooted in similar analyses as in, e.g., Rosenbaum and Rubin (1983b), Rosenbaum (2012), and Imbens (2003).

We concentrate on the model with secondary outcome and under the stochastic dominance assumption. Note that unconfoudedness states that the treatment is randomly assigned to the individuals conditional on $\mathbf{X}_i$, and this implies a balancing of any, observed and unobserved, pre-treatment variables between treatment groups. Principal strata are not affected by treatment assignment (Frangakis and Rubin, 2002) and can thus be regarded as pre-treatment variables. Therefore, under unconfoundedness, we have $P(G_i|\mathbf{X}_i, Z_i = 1) = P(G_i|\mathbf{X}_i, Z_i = 0)$, where, as before, G_i denotes the principal strata. And an implication of the failure of unconfoundedness is that $P(G_i|\mathbf{X}_i, Z_i = 1) \neq P(G_i|\mathbf{X}_i, Z_i = 0)$. We exploit this insight to assess the effects of interpretable unmeasured confounders by considering values of sensitivity parameters ξ_g ($g = EE, EN, NE$) for each of the strata in the model for the principal strata membership (the NN stratum will be set, without loss of generality, as the base category below). These sensitivity parameters will alter the equality above that must hold under unconfoundedness.

Importantly, including sensitivity parameters in the wage outcome model deserves attention in this setting because the consequences of an unmeasured confounder cannot be distinguished from the effect of the treatment since we are not imposing an exclusion restriction assumption (Schwartz et al., 2012). To explain, it is important to differentiate the assignment z in the definition of potential outcomes and the observed assignment Z, the values of which we denote by z_1 and z_2 respectively, so that the outcome model for the EE stratum can be formulated as:

$$E(\log[Y_{1,i}(z_1)]) = \mathbf{X}_i(\boldsymbol{\eta}_{EE,0} + \boldsymbol{\gamma} z_1 + \boldsymbol{\pi} z_2). \tag{2}$$

Here, the vector $\boldsymbol{\gamma}$ captures the treatment effect while the sensitivity vector $\boldsymbol{\pi}$ captures the unobserved confounding effect. However, since for each individual only the potential outcome corresponding to the observed treatment is observed, then $z_1 = z_2$ for any i, and thus only the sum $\boldsymbol{\gamma}+\boldsymbol{\pi}$ is identifiable. In other words, we cannot differentiate the two sets of parameter vectors $(\boldsymbol{\gamma}, \boldsymbol{\pi})$ and $(\boldsymbol{\gamma} + \boldsymbol{\nu}, \boldsymbol{\pi} - \boldsymbol{\nu})$ for any $\boldsymbol{\nu}$.

While we cannot separately identify the sensitivity parameters in the wage outcome model, this does not mean that we cannot completely assess the impact of selection into training (Z_i) on wages ($Y_{1,i}$). The sensitivity parameters ξ_g ($g = EE, EN, NE$) for each of the principal strata account for unobserved confounders that only jointly affect $\{Z_i$ and $S_i\}$. They also account for a portion of the confounding effect caused by the unobserved confounders that jointly affect $\{Z_i$, S_i, and $Y_{1,i}\}$. Our inability to include sensitivity parameters in the wage model ($\boldsymbol{\pi}$) prevents us from assessing the impact of (a) unobserved confounders that only jointly affect $\{Z_i$ and $Y_{1,i}\}$ plus (b) any "residual" confounding effect on wages of those unobserved confounders that jointly affect $\{Z_i$, S_i, and $Y_{1,i}\}$ that may remain once their confounding effect on the principal strata distribution between treatment groups has been taken into account. There are likely several instances in which this is not a serious shortcoming of the sensitivity analysis we employ, such as in our current empirical setting. More specifically, in our application, it is difficult to think of unobserved confounders that would be related only to language training (Z_i) and wages ($Y_{1,i}$) but not to employment (S_i).[15] In addition, it is sometimes possible to indirectly assess certain values of the sensitivity parameter $\boldsymbol{\pi}$ over the results from the available sensitivity analysis, as we illustrate below.

We now turn to the specification of the sensitivity analysis. For ease of exposition, consider probabilities that are not conditional on $\mathbf{X}_i$. 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 consider values of ξ_{EN} to increase the probability to be EN in the treatment arm relative to control: $P(G_i = EN|Z_i = 1) > P(G_i = EN|Z_i = 0)$. Presumably, EN are motivated to take the language training program (i.e., they may suspect they

15. This can be motivated by theoretical economic models of the labor market and the well-documented difficulty of finding variables related to employment but not related to wages (see the discussion and references mentioned in footnote 2.

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 consider values of ξ_{NE} to increase the probability to be NE in the treatment arm relative to control: $P(G_i = NE|Z_i = 1) > P(G_i = NE|Z_i = 0)$. One can think of the NE as individuals that, when treated, would raise their reservation wage (the minimum level of wage at which they are willing to work) 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 them raising their future reservation wage).

To set plausible values for the sensitivity parameters, we concentrate on values reflecting differences 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 4.[16] Note also that, for each Δ_g, the value of zero is consistent with the validity of the unconfoundedness assumption for stratum g.

To benchmark the relative importance of the departures from unconfoundedness considered in the sensitivity analysis, we consider the estimated marginal probabilities π_g for the local point ML3 in Table 4. For each Δ_g value, the following are the percentages, relative to the corresponding π_g in Table 4, 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.

We calculate, for each of the 27 scenarios under consideration, the implied sensitivity parameters ξ_g by applying 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 $(\Delta_{EE}, \Delta_{EN}, \Delta_{NE})$. Subsequently, the implied sensitivity parameters ξ_g are included in the model for the principal strata membership:

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

16. 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 $P(G_i = g|Z_i = 0) = \pi_{g,ML3} - (\Delta_g/2)$.

and the corresponding likelihood function is maximized.[17] We note that, while our approach to sensitivity analysis is tractable and easily interpretable, it represents an approximation in that we employ strata probabilities by treatment arm that are not conditional on X. Accounting for this conditioning would substantially complicate the procedure.[18]

The results of the sensitivity analysis are detailed in Appendix Tables C.4 to C.12 for all three local ML points detected in Table 4. Table 5 summarizes the range of estimated effects on employability and wages from the sensitivity analysis. Figure 1 visually summarizes the results for the effects on employability and wages for the local ML3 point in Table 4 (Appendix Tables C.10 to C.12). The sensitivity analysis yields similar results for each local ML point. The Appendix Tables show, 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 $(\Delta_{EN}, \Delta_{NE})$. In general, the values of $\hat{\pi}_{EE}$, $\hat{\pi}_{NE}$, $\hat{AT}(EE,0)$ and $\hat{AT}(NE,0)$ 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 $\hat{AT}.EE = \hat{AT}(EE,1) - \hat{AT}(EE,0)$, respectively, are due to the sensitivity observed in $\hat{\pi}_{EN}$ and $\hat{AT}(EE,1)$. 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 5 shows, for each local ML point, the minimum and maximum estimated effect on employability and wages, along with their standard errors (in parentheses) and the sensitivity scenario where they occur. It is observed that the effect on employment is relatively robust to the departures from unconfoundedness considered, particularly for local points ML2 and ML3. For these two local points, the minimum 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 4 was statistically insignificant.

The effect on wages for the always-employed is remarkably robust to the departures from unconfoundedness considered for all local ML points, since none of the estimated minimum or maximum effects are statistically significant, and they are all close to zero. Recall that our sensitivity analysis is not able to account for some unobserved confounders ($\boldsymbol{\pi}$ in equation (2)) since they are not distinguishable from the effect of the treatment due to the absence of the exclusion restriction assumption. In other words, the two sets of parameter vectors $(\boldsymbol{\gamma}, \boldsymbol{\pi})$ and $(\boldsymbol{\gamma} + \boldsymbol{\nu}, \boldsymbol{\pi} - \boldsymbol{\nu})$ cannot be distinguished for any $\boldsymbol{\nu}$. Nevertheless, based on the sensitivity analysis results, the remarkable stability around zero of the estimated effect on wages for the always-employed obtained over the several sensitivity scenarios is consistent with a negligible value for $\boldsymbol{\pi}$. Put another way, if $\boldsymbol{\pi}$ was

17. To decrease the computational burden in the estimation of the model 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.

18. 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 5: Sensitivity analysis, range of estimated effects on employability ($\hat{\pi}_{EN} - \hat{\pi}_{NE}$), and wages ($\hat{AT}.EE = \hat{AT}(EE, 1) - \hat{AT}(EE, 0)$).

		min		max	
		est. effect	$(\Delta_{EE}, \Delta_{EN}, \Delta_{NE})$	est. effect	$(\Delta_{EE}, \Delta_{EN}, \Delta_{NE})$
$\hat{\pi}_{EN} - \hat{\pi}_{NE}$	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 $(-0.15, 0, 0)$
$\hat{AT}.EE$	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	-0.16 (.36)	$(0, 0, 0)$	0.28 (.39)	$(-0.15, 0.05, 0.10)$

important, one would have to argue that the parameter $\boldsymbol{\gamma}$, which captures the treatment effect, varies over each sensitivity scenario in such a way so as to exactly balance to the same nonzero value for $\boldsymbol{\pi}$ ($\boldsymbol{\gamma} = \boldsymbol{\pi}$) each time, which seems very unlikely.

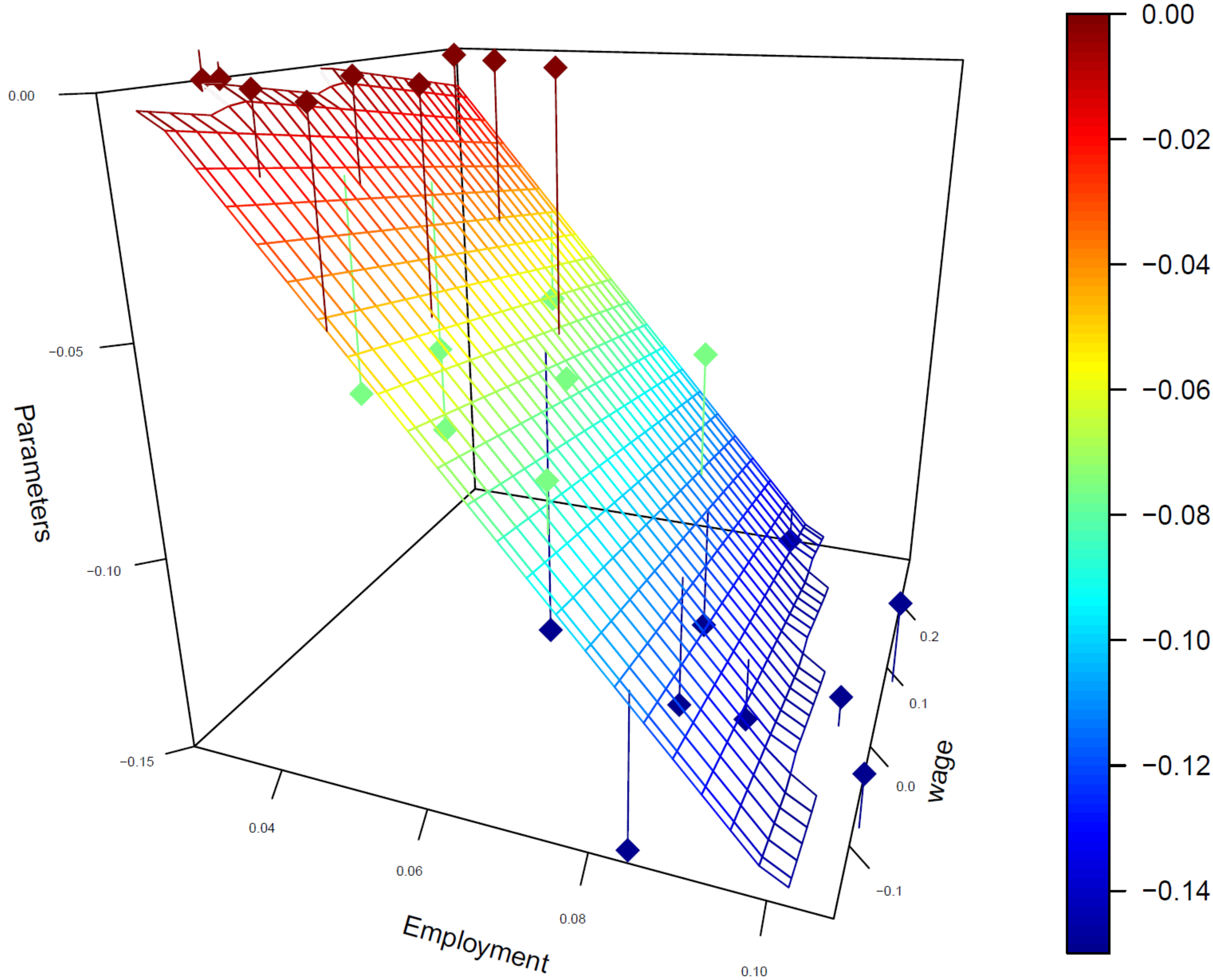
Figure 1 visually presents the sensitivity analysis results for the effects on employability and wages for the local ML3 point. It shows the different values of the employment (X axis) and wages (Y axis) estimated for $\Delta_{EE} = (0, -0.075, -0.15)$ (Z axis), conditional on the nine pairs of values for Δ_{EN} and Δ_{NE} equal to $(0, 0)$, $(0, 0.05)$, $(0, 0.10)$, $(0.05, 0)$, $(0.05, 0.05)$, $(0.05, 0.10)$, $(0.10, 0)$, $(0.10, 0.05)$, $(0.10, 0.10)$. The red points represent the estimated effects on the outcomes of interest conditional on $\Delta_{EE} = 0$, the green points the estimated effects conditional on $\Delta_{EE} = -0.075$, while the blue points are the estimated effects obtained conditional on $\Delta_{EE} = -0.15$. The range of the effects on the employment (minimum value 0.027, maximum value 0.107) and wage (minimum value -0.16, maximum value 0.28) (see Table 5) are included in the clusters of the red and blue points of Figure 1. Those are the absolute minimum and maximum effects of our estimated outcomes, respectively.

Summarizing the insights of the sensitivity analysis, with the exception of the local point ML1 for the employment effect, there is an acceptable robustness of the estimated effects on employability and wages shown in Table 4 to plausible departures from the crucial unconfoundedness assumption.

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 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 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

Figure 1: Sensitivity analysis



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.

For estimation, a normal mixture model is maximized using the EM algorithm within a frequentist 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 suggest that the language training for the unemployed in Luxembourg likely has a positive effect on 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 on the wages of those who would be employed regardless of training participation, we do not find evidence of a statistically significant effect. The estimated effects on both outcomes are shown to be largely robust to a set of plausible values for sensitivity parameters that model the presence of unmeasured confounders.

From a policy perspective, these findings suggest that the language training programs in Luxembourg likely have 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 potential explanations for these results. 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 comparatively 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 (Statec, 2012b). 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 condition for employment in the majority of jobs. It is of interest to examine in more detail the factors that may be behind the findings offered herein, particularly the relative importance of the second and third potential explanations.

FUNDING: 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.
CONFLICT OF INTEREST: the authors declare that they have no conflict of interest.

Acknowledgments

Part of this work was done when Andrea Mercatanti worked as Researcher at LISER and at the Bank of Italy. Michela Bia and Andrea Mercatanti acknowledge 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.

Appendix A. Summary Statistics

Table A.1 and A.2 present summary statistics for selected, individual and job, variables in the sample.

Table A.1: 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.

We provide a detailed description of some key variables included in the dataset with particular emphasis on the *employability score/level*. This variable strengthens the unconfoundedness assumption introduced in our identification strategy.

As described in the paper, the IGSS-ADEM dataset consists of administrative records derived from the global social security database in Luxembourg and a panel dataset on training programs collected by the Employment Agency in Luxembourg. The first data source includes detailed information on the trajectories of all workers employed in the country since the 1980s via their personal identification number. In the second data source the observation unit is represented by an "unemployment file", which corresponds to an unemployment spell. The ADEM dataset includes a rich set of information on the linked unemployed worker registered in ADEM from January 2007 to October 2011. For each unemployed individual it provides information on individual baseline characteristics (such as gender, age, nationality, civil status, health status and education) as well as individual labor market history (date of start of job, hourly wage, number of months worked in the last 12 months, activity sector, employment status previous to unemployment registration at ADEM, type of interventions assigned and date of start of the training program). Important to note: the ADEM dataset also includes information on a key variable, the so-called *score variable* defined to assess the employability level at baseline for each unemployed worker.

This "employability score" is a categorical variable with 5 dimensions/levels used by the caseworker to classify participants from very high (A) to very low (E) employability according to eight pre-treatment diagnostic dimensions: i) social setting; ii) employment trajectories; iii) health status; iv) psychological status; v) job prospects; vi) childcare; vii) mobility; and viii) job search behavior and evaluation of the unemployed profile. This variable, generally not available in similar studies, provides crucial information. It represents a proxy for generally unobservable individuals' characteristics, such as personality traits, expectations, skills and underlying motivation, which are strongly related to choice and job search behaviors of the subjects under study. Moreover, this employability score represents, for caseworkers, an objective way to steer individuals towards training programs, so it is a key determinant in the program assignment procedure. Therefore, we believe that the employability score is an important covariate that, together with the rich set of other pre-treatment variables, strengthens the plausibility of the unconfoundedness assumption.

Appendix B. Results: additional tables

Table B.3 presents the standard errors and the *p*-values obtained from the Delta method and the bootstrap procedure (with 100 replications) for the local ML points reported in Table 4 of the main document. These ML points are detected imposing the stochastic dominance restriction in the model with a secondary outcome. We take the ML parameter vector calculated from the original data as the starting value in applying the EM algorithm to each bootstrap sample. The key takeaway from this table is that there is a complete agreement between the conclusions achieved with the standard errors computed using the Delta method and the bootstrap.

Appendix C. Sensitivity analysis: additional tables

Table C.4 - Table C.12 present the results from the sensitivity analysis for the local points (ML1, ML2 and ML3) reported in Table 4 of the main document. For a given value of Δ_{EE}, each of these tables reports: the estimated strata probabilities, the estimated average wages, and the effects on employability and wages, that correspond to each value of the pair (Δ_{EN} and Δ_{NE})). See the main text for details about the sensitivity analysis.

References

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

J. Angrist and A. Krueger. *Empirical Strategies in Labor Economics*. Handbook of Labor Economics, O. Ashen-felter and D. Card, ed.s. Amsterdam: North-Holland, 1999.

J. Angrist, G. Imbens, and D. Rubin. Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, 91:444–445, 1996.

G. Blanco, A. Flores-Lagunes, and Flores C.A. Bounds on the quantile treatment effects of job corps on wages. *Journal of Human Resources*, 483:659–701, 2013.

R. Blundell, A. Gosling, H. Ichimura, and C. Meguir. Changes in the distribution of male and female wages accounting for employment composition using bounds. *Econometrica*, 75:387–399, 2007.

J. Brosius and M. Zanardelli. Evaluation de l' efficacité des mesures de formation destinées aux chómeurs. *LISER, Technical Report, Ministry of Labour, Employment, Social Economy and Economic Solidarity of Luxembourg*, 2012.

M. Buot and D. Richards. Counting and locating the solutions of polynomial systems of maximum likelihood equations. *Journal of Symbolic Computation*, 41:234–244, 2006.

F. Catanese, S. Hosten, A. Khetan, and B. Sturmfels. The maximum likelihood degree. *American Journal of Mathematics*, 128:671–697, 2006.

G. Ciuperca, A. Ridolfi, and J. Idier. Penalized maximum likelihood estimator for normal mixtures. *Scandinavian Journal of Statistics*, 30:45–59, 2003.

J. Clausen, H. Hummelgaard, L. Husted, H. Blume, and M. Rosholm. The impact of the introduction programme on the labour-market integration. *Copenhagen: Institute for Local Studies*, 30: 45–59, 2006.

H. Cramér. *Mathematical Methods of Statistics*. Princeton, NJ: Princeton University Press, 1946.

A. Dempster, N. Laird, and D. Rubin. Maximum likelihood from incomplete data via the em algorithm (with discussion). *Journal of the Royal Statistical Society - Series B*, 39:1–38, 1977.

A. Donado. Foreign languages and their impact on unemployment. *Labour*, 31:265–287, 2017.

C. Dustmann and F. Fabbri. Language proficiency and labour market performance of immigrants in the uk. *The Economic Journal*, 113:695–717, 2003.

European Commission. European social fund in luxembourg (2015): Report of the european commission. 2015.

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

P. Frumento, F. Mealli, B. Pacini, and D. Rubin. 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, 2012.

L. Gan and J. Jiang. A test for a global maximum. *Journal of the American Statistical Association*, 94:847–854, 1999.

M. Gerfin and M. Lechner. A microeconometric evaluation of the active labour market policy in switzerland. *The Economic Journal*, 112:854–893, 2002.

V. Ginsburgh and J. Prieto-Rodriguez. Returns to foreign languages of native workers in the eu. *Industrial and Labor Relations Review*, 64:599–618, 2011.

R.J. Hathaway. A constrained formulation of maximum-likelihood estimation for normal mixture distributions. *The Annals of Statistics*, 13:795–800, 1985.

J. Heckman. Sample selection bias as a specification error. *Econometrica*, 47:153–161, 1979.

J. Heckman, R. LaLonde, and J. Smith. *The Economics and Econometrics of Active Labor Market Programs*. in Handbook of Labor Economics Vol.III, O. Ashenfelter, and D. Card, ed.s, pp. 1865–2097. Elsevier, Amsterdam, 1999.

G. Imbens. Sensitivity to exogeneity assumptions in program evaluation. *The American Economic Review*, 93:126–132, 2003.

W. Imbens and J. Angrist. Identification and estimation of local average treatment effects. *Econometrica*, 62:467–475, 1994.

I. Isphording. Returns to foreign language skills of immigrants in spain. *Labour*, 27:443–461, 2013.

I. Isphording. Language and labor market success. *IZA working paper DP No. 8572*, 2014.

I. Isphording and S. Otten. The costs of babylon—linguistic distance in applied economics. *Review of International Economics*, 21:354–369, 2013.

M. Killingsworth. *Labor Supply*. Cambridge Surveys of Economic Literature. Cambridge University Press, 1983.

P. Kuhn and F. Lozano. The expanding workweek? understanding trends in long work hours among u.s. men, 1979-2006. *Journal of Labor Economics*, 26:311–343, 2008.

M. Lechner. Earnings and employment effects of continuous off-the-job training in east germany after unification. *Journal of Business and Economic Statistics*, 17:74–90, 1999.

M. Lechner and B. Melly. Partial identification of wage effects of training programs. *Unpublished*, 2010.

M. Lechner and C. Wunsch. Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *Review of Economic Studies*, 76:1071–1102, 2009.

M. Lechner and C. Wunsch. Sensitivity of matching-based program evaluations to the availability of control variables. *Labour Economics*, 2013.

E. Lehmann and G. Casella. *Theory of point estimation*. 2nd edn. New York: Springer-Verlag, 1998.

J. Lohmann. Do language barriers affect trade? *Economics Letters*, 110:159–162, 2011.

A. Mattei, F. Li, and F. Mealli. 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, 2013.

M. McHugh and A. Challinor. Improving immigrants' employment prospects through work-focused language instruction. *Migration Policy Institute (US)*, 2011.

G. McLachlan and D. Peel. *Finite Mixture Models*. New York: Wiley, 2000.

F. Mealli and B. Pacini. Using secondary outcomes to sharpen inference in randomized experiments with noncompliance. *Journal of the American Statistical Association*, 108:1120–1131, 2013.

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

A. Mercatanti, F. Li, and F. Mealli. Improving inference of gaussian mixtures using auxiliary variables. *Statistical Analysis and Data Mining*, 8:34–48, 2015.

OECD. *Jobs for Immigrants*. Volume 1: Labour Market Integration in Australia, Denmark, Germany and Sweden. Paris: OECD, 2007.

OECD. *Economic Surveys: Luxembourg (2010)*. Paris: OECD, 2010.

OECD. *Economic Surveys: Luxembourg (2010)*. Paris: OECD, 2012.

P. Rosenbaum. *Observational studies*. Second Edition. Springer-Verlag: New York, 2012.

P. Rosenbaum and D. Rubin. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 8:41–55, 1983a.

P. Rosenbaum and D. Rubin. Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. *Journal of the Royal Statistical Society - Series B*, 45: 212–218, 1983b.

D. Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66:688–701, 1974.

D. Rubin. Bayesian inference for causal effects. *Annals of Statistics*, 6:34–58, 1978.

D. Rubin. Discussion of "randomization analysis of experimental data: the fisher randomization test" by d. basu. *Journal of the American Statistical Association*, 75:91–593, 1980.

S. Schwartz, F. Li, and J. Reiter. Sensitivity analysis for unmeasured confounding in principal stratification. *Statistics in Medicine*, 106:1331–1344, 2012.

S. Seaman, J. Galati, D. Jackson, and J. Carlin. What is meant by "missing at random"? *Statistics in Medicine*, 28:257–268, 2013.

C. Small, J. Wang, and Z. Yang. Eliminating multiple root problems inestimation. *Statistical Science*, 15:313–332, 2000.

Statec. Population by nationality. *http://www.statistiques.public.lu/en/news/population/population/2012/08/2012082* 2012a.

Statec. Bulletin du statec. *Statec, Luxembourg*, 2012b.

Statec. Luxembourg un demi-siècle de constantes et de variables. *ISBN : 978-2-87988-112-6, Statec, Luxembourg*, 2013.

Statec. Luxembourg en chiffres 2014. *http://www.statistiques.public.lu/catalogue-publications/luxembourgenchiffres/ luxembourg-chiffres.pdf*, 2014a.

Statec. Note de conjoncture 2014–2. *http://www.statistiques.public.lu/catalogue-publications/noteconjoncture/2014/PDFNDC-02-14.pdf*, 2014b.

A. Wald. Note on the consistency of the maximum likelihood estimate. *The Annals of Mathematical Statistics*, 20:595–560, 1949.

D. Williams. Multiple language usage and earnings in western europe. *International Journal of Manpower*, 32:372–393, 2011.

J. Zhang, D. Rubin, and F. Mealli. *Evaluating the Effects of Job Training Programs onWages Through Principal Stratification,*. D. Millimet, J. Smith and E. Vytlacil (eds.) Modelling and Evaluating Treatment Effects in Econometrics, Vol. 21 (Amsterdam: Elsevier), 2008.

J. Zhang, D. Rubin, and Mealli F. Journal of the american statistical association. *Likelihood-based analysis of causal effects of job-training programs using principal stratificatiol Association*, 104: 166–176, 2009.

Table A.2: 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

Table B.3: Local ML points detected with secondary outcome and the stochastic dominance restriction: standard errors and *p*-value from the Delta Method (D.M.) and the bootstrap (with 100 replications).

		Standard error		*p*-value	
		D.M.	Bootstrap	D.M:	Bootstrap
ML1					
$\hat{\pi}_{EE}$	0.375	(.003)	(.003)	(.000)	(.000)
$\hat{\pi}_{EN}$	0.123	(.009)	(.012)	(.000)	(.000)
$\hat{\pi}_{NE}$	0.115	(.002)	(.002)	(.000)	(.000)
$\hat{\pi}_{NN}$	0.386	(.009)	(.012)	(.000)	(.000)
Est. effect on employment $\hat{\pi}_{EN} - \hat{\pi}_{NE}$	0.008	(.009)	(.012)	(.374)	(.505)
$\hat{AT}(EE, 1)$	14.80	(.39)	(.37)	(.000)	(.000)
$\hat{AT}(EE, 0)$	15.05	(.07)	(.08)	(.000)	(.000)
$\hat{AT}(EN, 1)$	14.21	(.75)	(.60)	(.000)	(.000)
$\hat{AT}(NE, 0)$	13.47	(.13)	(.10)	(.000)	(.000)
Est. effect on hourly wages for treated EE	−0.25	(.40)	(.37)	(.532)	(.500)
ML2					
$\hat{\pi}_{EE}$	0.375	(.003)	(.003)	(.000)	(.000)
$\hat{\pi}_{EN}$	0.241	(.010)	(.011)	(.000)	(.000)
$\hat{\pi}_{NE}$	0.114	(.002)	(.002)	(.000)	(.000)
$\hat{\pi}_{NN}$	0.269	(.011)	(.011)	(.000)	(.000)
Est. effect on employment $\hat{\pi}_{EN} - \hat{\pi}_{NE}$	0.127	(.011)	(.012)	(.000)	(.000)
$\hat{AT}(EE, 1)$	14.73	(.39)	(.33)	(.000)	(.000)
$\hat{AT}(EE, 0)$	14.99	(.07)	(.09)	(.000)	(.000)
$\hat{AT}(EN, 1)$	11.91	(.82)	(.59)	(.000)	(.000)
$\hat{AT}(NE, 0)$	13.38	(.13)	(.10)	(.000)	(.000)
Est. effect on hourly wages for treated EE	−0.24	(.39)	(.34)	(.538)	(.480)
ML3					
$\hat{\pi}_{EE}$	0.377	(.003)	(.003)	(.000)	(.000)
$\hat{\pi}_{EN}$	0.166	(.010)	(.019)	(.000)	(.000)
$\hat{\pi}_{NE}$	0.114	(.002)	(.002)	(.000)	(.000)
$\hat{\pi}_{NN}$	0.342	(.010)	(.018)	(.000)	(.000)
Est. effect on employment $\hat{\pi}_{EN} - \hat{\pi}_{NE}$	0.052	(.010)	(.019)	(.000)	(.006)
$\hat{AT}(EE, 1)$	14.84	(.36)	(.42)	(.000)	(.000)
$\hat{AT}(EE, 0)$	15.00	(.07)	(.09)	(.000)	(.000)
$\hat{AT}(EN, 1)$	10.23	(1.11)	(1.14)	(.000)	(.000)
$\hat{AT}(NE, 0)$	13.38	(.13)	(.10)	(.000)	(.000)
Est. effect on hourly wages for treated EE	−0.16	(.36)	(.44)	(0.657)	(0.716)

Table C.4: Sensitivity analysis, ML1, $\Delta_{EE} = 0$

	$\hat{\pi}_{EE}$	$\hat{\pi}_{EN}$	$\hat{\pi}_{NE}$	$\hat{\pi}_{NN}$	$\hat{\pi}_{EN} - \hat{\pi}_{NE}$	$\hat{AT}(EE,1)$	$\hat{AT}(EE,0)$	$\hat{AT}.EE$	$\hat{AT}(EN,1)$	$\hat{AT}(NE,0)$
$\Delta_{EN} = 0$	.375	.123	.115	.386	.008	14.80	15.05	-.25	14.21	13.47
$\Delta_{NE} = 0$	(.003)	(.009)	(.002)	(.009)	(.009)	(.39)	(.07)	(.40)	(.75)	(.13)
$\Delta_{EN} = 0$	.375	.131	.116	.377	.015	14.85	15.05	-.20	14.29	13.49
$\Delta_{NE} = 0.05$	(.003)	(.009)	(.002)	(.009)	(.008)	(.39)	(.07)	(.40)	(.75)	(.13)
$\Delta_{EN} = 0$	.375	.145	.118	.361	.027	14.94	15.04	-.10	14.33	13.49
$\Delta_{NE} = 0.10$	(.003)	(.008)	(.002)	(.008)	(.008)	(.40)	(.07)	(.40)	(.75)	(.13)
$\Delta_{EN} = 0.05$	.375	.103	.115	.406	-.012	14.74	15.06	-.32	14.21	13.47
$\Delta_{NE} = 0$	(.003)	(.009)	(.002)	(.009)	(.009)	(.39)	(.07)	(.40)	(.75)	(.13)
$\Delta_{EN} = 0.05$	.375	.111	.116	.396	-.005	14.79	15.05	-.26	14.25	13.50
$\Delta_{NE} = 0.05$	(.003)	(.009)	(.002)	(.009)	(.009)	(.40)	(.07)	(.40)	(.76)	(.13)
$\Delta_{EN} = 0.05$	.375	.124	.118	.382	.006	14.88	15.05	-.17	14.35	13.52
$\Delta_{NE} = 0.10$	(.003)	(.008)	(.002)	(.008)	(.008)	(.40)	(.07)	(.41)	(.76)	(.13)
$\Delta_{EN} = 0.10$	.375	.086	.115	.423	-.029	14.69	15.06	-.37	14.21	13.48
$\Delta_{NE} = 0$	(.003)	(.009)	(.002)	(.009)	(.009)	(.40)	(.07)	(.40)	(.75)	(.13)
$\Delta_{EN} = 0.10$	.375	.091	.116	.417	-.025	14.74	15.06	-.32	14.30	13.50
$\Delta_{NE} = 0.05$	(.003)	(.009)	(.002)	(.009)	(.009)	(.41)	(.07)	(.41)	(.76)	(.13)
$\Delta_{EN} = 0.10$	.375	.103	.118	.402	-.015	14.82	15.06	-.24	14.35	13.52
$\Delta_{NE} = 0.10$	(.003)	(.008)	(.002)	(.008)	(.008)	(.41)	(.07)	(.41)	(.76)	(.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 C.5: Sensitivity analysis, ML1, $\Delta_{EE} = -0.075$

	$\hat{\pi}_{EE}$	$\hat{\pi}_{EN}$	$\hat{\pi}_{NE}$	$\hat{\pi}_{NN}$	$\hat{\pi}_{EN} - \hat{\pi}_{NE}$	$\hat{AT}(EE,1)$	$\hat{AT}(EE,0)$	$\hat{AT}.EE$	$\hat{AT}(EN,1)$	$\hat{AT}(NE,0)$
$\Delta_{EN} = 0$	.375	.131	.115	.379	.016	14.86	15.05	-.20	14.22	13.46
$\Delta_{NE} = 0$	(.003)	(.009)	(.002)	(.009)	(.009)	(.39)	(.07)	(.39)	(.74)	(.13)
$\Delta_{EN} = 0$	.375	.137	.116	.372	.021	14.91	15.05	-.14	14.26	13.48
$\Delta_{NE} = 0.05$	(.003)	(.009)	(.002)	(.009)	(.009)	(.39)	(.07)	(.39)	(.74)	(.13)
$\Delta_{EN} = 0$	.375	.149	.117	.358	.032	15.00	15.05	-.05	14.35	13.50
$\Delta_{NE} = 0.10$	(.003)	(.008)	(.002)	(.008)	(.008)	(.39)	(.07)	(.40)	(.75)	(.13)
$\Delta_{EN} = 0.05$	.375	.110	.115	.399	-.005	14.79	15.06	-.27	14.16	13.47
$\Delta_{NE} = 0$	(.003)	(.009)	(.002)	(.009)	(.009)	(.39)	(.07)	(.39)	(.75)	(.13)
$\Delta_{EN} = 0.05$	.375	.115	.116	.393	-.001	14.85	15.06	-.21	14.27	13.48
$\Delta_{NE} = 0.05$	(.003)	(.008)	(.002)	(.008)	(.008)	(.39)	(.07)	(.40)	(.76)	(.13)
$\Delta_{EN} = 0.05$	.375	.126	.117	.381	.009	14.93	15.05	-.12	14.32	13.50
$\Delta_{NE} = 0.10$	(.003)	(.008)	(.002)	(.008)	(.008)	(.40)	(.07)	(.40)	(.76)	(.13)
$\Delta_{EN} = 0.10$	.375	.090	.115	.420	-.025	14.72	15.07	-.35	14.24	13.48
$\Delta_{NE} = 0$	(.003)	(.008)	(.002)	(.009)	(.009)	(.39)	(.07)	(.40)	(.75)	(.13)
$\Delta_{EN} = 0.10$	.375	.095	.116	.417	-.021	14.78	15.06	-.28	14.28	13.49
$\Delta_{NE} = 0.05$	(.003)	(.008)	(.002)	(.008)	(.008)	(.40)	(.07)	(.41)	(.76)	(.13)
$\Delta_{EN} = 0.10$	.375	.104	.117	.403	-.013	14.87	15.06	-.19	14.36	13.50
$\Delta_{NE} = 0.10$	(.003)	(.007)	(.002)	(.008)	(.008)	(.40)	(.07)	(.41)	(.76)	(.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 C.6: Sensitivity analysis, ML1, $\Delta_{EE} = -0.15$

	$\hat{\pi}_{EE}$	$\hat{\pi}_{EN}$	$\hat{\pi}_{NE}$	$\hat{\pi}_{NN}$	$\hat{\pi}_{EN} - \hat{\pi}_{NE}$	$\hat{AT}(EE,1)$	$\hat{AT}(EE,0)$	$\hat{AT}.EE$	$\hat{AT}(EN,1)$	$\hat{AT}(NE,0)$
$\Delta_{EN} = 0$	.375	.107	.116	.402	-.009	14.91	15.07	-.16	14.35	13.46
$\Delta_{NE} = 0$	(.002)	(.008)	(.002)	(.008)	(.008)	(.39)	(.07)	(.39)	(.74)	(.13)
$\Delta_{EN} = 0$	.375	.145	.116	.364	.029	14.98	15.05	-.07	14.35	13.46
$\Delta_{NE} = 0.05$	(.002)	(.008)	(.002)	(.008)	(.008)	(.39)	(.07)	(.39)	(.74)	(.13)
$\Delta_{EN} = 0$	.375	.156	.117	.352	.039	15.07	15.05	.02	14.36	13.47
$\Delta_{NE} = 0.10$	(.002)	(.008)	(.002)	(.008)	(.008)	(.40)	(.07)	(.40)	(.75)	(.13)
$\Delta_{EN} = 0.05$	.375	.118	.114	.392	.004	14.84	15.06	-.22	14.25	13.46
$\Delta_{NE} = 0$	(.002)	(.008)	(.002)	(.008)	(.008)	(.39)	(.07)	(.39)	(.74)	(.13)
$\Delta_{EN} = 0.05$	.375	.121	.116	.388	.005	14.90	15.06	-.16	14.27	13.47
$\Delta_{NE} = 0.05$	(.002)	(.008)	(.002)	(.008)	(.008)	(.39)	(.07)	(.39)	(.74)	(.13)
$\Delta_{EN} = 0.05$	.375	.130	.117	.377	.013	14.99	15.06	-.07	14.38	13.47
$\Delta_{NE} = 0.10$	(.002)	(.008)	(.002)	(.008)	(.008)	(.40)	(.07)	(.40)	(.75)	(.13)
$\Delta_{EN} = 0.10$	.375	.096	.114	.414	-.018	14.77	15.07	-.30	14.26	13.47
$\Delta_{NE} = 0$	(.002)	(.008)	(.002)	(.008)	(.008)	(.39)	(.07)	(.39)	(.74)	(.13)
$\Delta_{EN} = 0.10$	.375	.099	.116	.410	-.017	14.83	15.07	-.24	14.33	13.48
$\Delta_{NE} = 0.05$	(.002)	(.008)	(.002)	(.008)	(.008)	(.40)	(.07)	(.40)	(.75)	(.13)
$\Delta_{EN} = 0.10$	.375	.107	.117	.400	-.010	14.92	15.06	-.14	14.39	13.48
$\Delta_{NE} = 0.10$	(.002)	(.008)	(.002)	(.008)	(.008)	(.40)	(.07)	(.40)	(.75)	(.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 C.7: Sensitivity analysis, ML2, $\Delta_{EE} = 0$

	$\hat{\pi}_{EE}$	$\hat{\pi}_{EN}$	$\hat{\pi}_{NE}$	$\hat{\pi}_{NN}$	$\hat{\pi}_{EN} - \hat{\pi}_{NE}$	$\hat{AT}(EE,1)$	$\hat{AT}(EE,0)$	$\hat{AT}.EE$	$\hat{AT}(EN,1)$	$\hat{AT}(NE,0)$
$\Delta_{EN} = 0$	.375	.241	.114	.269	.127	14.73	14.99	-.26	11.91	13.38
$\Delta_{NE} = 0$	(.003)	(.010)	(.002)	(.011)	(.011)	(.39)	(.007)	(.39)	(.82)	(.13)
$\Delta_{EN} = 0$	.375	.246	.116	.262	.130	14.83	14.99	-.16	11.95	13.38
$\Delta_{NE} = 0.05$	(.003)	(.010)	(.002)	(.011)	(.011)	(.39)	(.007)	(.39)	(.81)	(.13)
$\Delta_{EN} = 0$	.375	.256	.117	.251	.139	14.98	14.98	.00	12.03	13.38
$\Delta_{NE} = 0.10$	(.003)	(.010)	(.002)	(.011)	(.011)	(.39)	(.007)	(.39)	(.81)	(.13)
$\Delta_{EN} = 0.05$	.375	.220	.115	.290	.105	14.80	14.99	-.19	11.89	13.38
$\Delta_{NE} = 0$	(.003)	(.010)	(.002)	(.011)	(.011)	(.39)	(.007)	(.39)	(.82)	(.13)
$\Delta_{EN} = 0.05$	.375	.224	.116	.284	.108	14.90	14.98	-.08	11.96	13.38
$\Delta_{NE} = 0.05$	(.003)	(.010)	(.002)	(.011)	(.011)	(.39)	(.007)	(.39)	(.82)	(.13)
$\Delta_{EN} = 0.05$	.375	.234	.117	.273	.117	15.05	14.98	.07	12.00	13.39
$\Delta_{NE} = 0.10$	(.003)	(.010)	(.002)	(.011)	(.011)	(.39)	(.007)	(.39)	(.82)	(.13)
$\Delta_{EN} = 0.10$	.375	.197	.115	.313	.082	14.87	14.99	-.11	11.85	13.38
$\Delta_{NE} = 0$	(.003)	(.010)	(.002)	(.010)	(.010)	(.40)	(.007)	(.40)	(.81)	(.13)
$\Delta_{EN} = 0.10$	.375	.201	.116	.307	.085	14.97	14.98	.01	11.98	13.39
$\Delta_{NE} = 0.05$	(.003)	(.010)	(.002)	(.010)	(.010)	(.40)	(.007)	(.40)	(.81)	(.13)
$\Delta_{EN} = 0.10$	.375	.211	.118	.296	.093	15.11	14.98	.13	12.05	13.39
$\Delta_{NE} = 0.10$	(.003)	(.010)	(.002)	(.010)	(.010)	(.40)	(.07)	(.40)	(.80)	(.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 C.8: Sensitivity analysis, ML2, $\Delta_{EE} = -0.075$

	$\hat{\pi}_{EE}$	$\hat{\pi}_{EN}$	$\hat{\pi}_{NE}$	$\hat{\pi}_{NN}$	$\hat{\pi}_{EN} - \hat{\pi}_{NE}$	$\hat{AT}(EE,1)$	$\hat{AT}(EE,0)$	$\hat{AT}.EE$	$\hat{AT}(EN,1)$	$\hat{AT}(NE,0)$
$\Delta_{EN} = 0$	.375	.251	.114	.260	.137	14.71	14.98	-.27	11.93	13.38
$\Delta_{NE} = 0$	(.003)	(.010)	(.002)	(.010)	(.011)	(.39)	(.07)	(.39)	(.81)	(.13)
$\Delta_{EN} = 0$	.375	.253	.115	.256	.138	14.82	14.99	-.17	11.98	13.38
$\Delta_{NE} = 0.05$	(.003)	(.011)	(.002)	(.011)	(.11)	(.39)	(.07)	(.39)	(.81)	(.13)
$\Delta_{EN} = 0$	.375	.262	.117	.246	.145	14.97	14.99	-.02	12.05	13.37
$\Delta_{NE} = 0.10$	(.003)	(.011)	(.002)	(.011)	(.11)	(.39)	(.07)	(.40)	(.82)	(.13)
$\Delta_{EN} = 0.05$	.375	.227	.114	.283	.113	14.79	14.97	-.18	11.90	13.38
$\Delta_{NE} = 0$	(.003)	(.011)	(.002)	(.011)	(.11)	(.39)	(.07)	(.39)	(.81)	(.13)
$\Delta_{EN} = 0.05$	.375	.230	.115	.279	.115	14.90	14.99	-.09	11.96	13.38
$\Delta_{NE} = 0.05$	(.003)	(.011)	(.002)	(.011)	(.11)	(.39)	(.07)	(.39)	(.81)	(.13)
$\Delta_{EN} = 0.05$	.375	.238	.117	.269	.121	15.04	14.98	.06	12.01	13.37
$\Delta_{NE} = 0.10$	(.003)	(.011)	(.002)	(.011)	(.11)	(.39)	(.07)	(.40)	(.82)	(.13)
$\Delta_{EN} = 0.10$	.375	.203	.114	.308	.089	14.86	15.00	-.14	11.87	13.38
$\Delta_{NE} = 0$	(.003)	(.011)	(.002)	(.011)	(.11)	(.39)	(.07)	(.39)	(.81)	(.13)
$\Delta_{EN} = 0.10$	.375	.206	.116	.303	.090	14.97	14.99	-.02	11.91	13.37
$\Delta_{NE} = 0.05$	(.003)	(.011)	(.002)	(.011)	(.11)	(.39)	(.07)	(.40)	(.82)	(.13)
$\Delta_{EN} = 0.10$	.375	.214	.117	.293	.097	15.12	14.98	.14	11.96	13.37
$\Delta_{NE} = 0.10$	(.003)	(.011)	(.002)	(.011)	(.010)	(.39)	(.07)	(.40)	(.82)	(.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 C.9: Sensitivity analysis, ML2, $\Delta_{EE} = -0.15$

	$\hat{\pi}_{EE}$	$\hat{\pi}_{EN}$	$\hat{\pi}_{NE}$	$\hat{\pi}_{NN}$	$\hat{\pi}_{EN} - \hat{\pi}_{NE}$	$\hat{AT}(EE,1)$	$\hat{AT}(EE,0)$	$\hat{AT}.EE$	$\hat{AT}(EN,1)$	$\hat{AT}(NE,0)$
$\Delta_{EN} = 0$	.375	.221	.115	.288	.106	14.99	15.00	-.01	11.94	13.36
$\Delta_{NE} = 0$	(.003)	(.010)	(.002)	(.010)	(.010)	(.39)	(.07)	(.39)	(.79)	(.13)
$\Delta_{EN} = 0$	.375	.263	.115	.246	.148	14.78	15.00	-.23	11.99	13.36
$\Delta_{NE} = 0.05$	(.003)	(.010)	(.002)	(.010)	(.010)	(.39)	(.07)	(.39)	(.79)	(.13)
$\Delta_{EN} = 0$	.375	.269	.116	.239	.153	14.94	14.99	-.05	12.06	13.36
$\Delta_{NE} = 0.10$	(.003)	(.010)	(.002)	(.010)	(.010)	(.39)	(.07)	(.39)	(.80)	(.13)
$\Delta_{EN} = 0.05$	.375	.236	.114	.274	.122	14.76	15.00	-.24	11.92	13.36
$\Delta_{NE} = 0$	(.003)	(.010)	(.002)	(.010)	(.010)	(.39)	(.07)	(.39)	(.79)	(.13)
$\Delta_{EN} = 0.05$	.375	.238	.115	.272	.123	14.87	15.00	-.13	11.99	13.36
$\Delta_{NE} = 0.05$	(.003)	(.010)	(.002)	(.010)	(.010)	(.39)	(.07)	(.39)	(.80)	(.13)
$\Delta_{EN} = 0.05$	.375	.244	.116	.264	.128	15.03	14.99	.04	12.02	13.36
$\Delta_{NE} = 0.10$	(.003)	(.010)	(.002)	(.010)	(.010)	(.39)	(.07)	(.39)	(.80)	(.13)
$\Delta_{EN} = 0.10$	.375	.210	.114	.301	.096	14.85	15.00	-.15	11.88	13.36
$\Delta_{NE} = 0$	(.003)	(.010)	(.002)	(.010)	(.010)	(.39)	(.07)	(.39)	(.80)	(.13)
$\Delta_{EN} = 0.10$	.375	.212	.115	.298	.097	14.96	14.99	-.03	11.93	13.36
$\Delta_{NE} = 0.05$	(.003)	(.010)	(.002)	(.010)	(.010)	(.39)	(.07)	(.39)	(.81)	(.13)
$\Delta_{EN} = 0.10$	.375	.219	.116	289	.103	15.11	14.99	.12	11.97	13.36
$\Delta_{NE} = 0.10$	(.003)	(.010)	(.002)	(.010)	(.010)	(.39)	(.07)	(.39)	(.81)	(.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 C.10: Sensitivity analysis estimates for ML3; $\Delta_{EE} = 0$

Δ_{EE}	$\hat{\pi}_{EE}$	$\hat{\pi}_{EN}$	$\hat{\pi}_{NE}$	$\hat{\pi}_{NN}$	$\hat{\pi}_{EN} - \hat{\pi}_{NE}$	$\hat{AT}(EE,1)$	$\hat{AT}(EE,0)$	$\hat{AT}.EE$	$\hat{AT}(EN,1)$	$\hat{AT}(NE,0)$
$\Delta_{EN} = 0$	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 C.11: Sensitivity analysis estimates for ML3; $\Delta_{EE} = -0.075$.

Δ_{EE}	$\hat{\pi}_{EE}$	$\hat{\pi}_{EN}$	$\hat{\pi}_{NE}$	$\hat{\pi}_{NN}$	$\hat{\pi}_{EN} - \hat{\pi}_{NE}$	$\hat{AT}(EE,1)$	$\hat{AT}(EE,0)$	$\hat{AT}.EE$	$\hat{AT}(EN,1)$	$\hat{AT}(NE,0)$
$\Delta_{EN} = 0$	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	0.21	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 C.12: Sensitivity analysis estimates for ML3; $\Delta_{EE} = -0.15$.

Δ_{EE}	$\hat{\pi}_{EE}$	$\hat{\pi}_{EN}$	$\hat{\pi}_{NE}$	$\hat{\pi}_{NN}$	$\hat{\pi}_{EN} - \hat{\pi}_{NE}$	$\hat{AT}(EE,1)$	$\hat{AT}(EE,0)$	$\hat{AT}.EE$	$\hat{AT}(EN,1)$	$\hat{AT}(NE,0)$
$\Delta_{EN} = 0$	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)

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