Pre-analysis plan – La Bonne Alternance Experiment

August 29, 2022

Clément Brébion[1] (CBS) and Simon Briole[2] (CEEM - PSE)

Introduction

This document outlines the analysis to be carried out using data from the *LBA* experiment. Its purpose is to outline the specifications to be used in the analysis, limiting concerns about specification search.

The goal of the study is to analyze the effects of an internet-based recommender system that identifies establishments likely to hire apprentices in a geographical area of interest. Such a system is expected to improve the targeting of unsolicited applications, thereby enhancing the quality and number of matches between firms and candidates to apprenticeship programs.

The system we analyze is a French website, *La Bonne Alternance* (LBA), rolled out by the French Public Employment Service (*Pôle Emploi*) in 2018. It is receiving in 2021 about 1.5 million visits per year and displays about 500.000 plants. It has received sizable attention in the public debate because of both its audience and the difficulties to foster high-quality apprenticeships in France.

In terms of methods, the research project builds on a randomized experiment that inhibits some of the plants normally displayed on the website. The experiment started in September 2021 and will run until the end of September 2022. The data analysis will start in October-November 2022, using administrative databases on the French Ministry of Labor (DARES) premises.

The plan is outlined as follows: Section 1 discusses the program objectives, outcomes and measurement. Section 2 describes the study population and the randomization process. Section 3 describes the internal validity checks we will perform. Section 4 enumerates the hypotheses we will test regarding the effects of the treatment in the whole sample. Section 5 describes the heterogeneity analysis we will perform.

I. Program objectives, outcomes and measurement

The program

The LBA website is jointly managed by an inter-ministerial mission for apprenticeship training ("La mission apprentissage") and Pôle Emploi (PE), the French Public Employment Service. It aims at improving the number and the quality of apprenticeship contracts by fostering and improving the outcome of unsolicited applications.

[1] Email address: cbr.msc@cbs.dk
[2] Email address: simon.briole@psemail.eu

On the website, one can find a list of establishments – with their contact details – that are likely to hire apprentices in the selected occupation, at the preferred level of studies and in the geographical area of interest.

The website relies on a machine-learning algorithm created by PE that predicts : (i) hiring on apprenticeship contracts at the plant x occupation level ; (ii) hiring on permanent contracts or fixed-term contracts of more than one month at the plant x occupation level.[3] The algorithm follows three steps. First, a regression predicts the number of hirings into apprenticeship/standard employment contracts to expect from each plant. Second, PE weights each of these two predictions using an index that accounts for the distribution of hirings in standard employment across occupations in each sector. The two resulting (weighted) scores at the plant x occupation level are finally compared with an arbitrary threshold. Establishments are displayed on the website for an occupation if at least one of the two scores at the plant x occupation level exceeds this threshold.

Outcomes

This study aims at evaluating the impact of the LBA website on plants' training behaviour on the short and medium run. In total, about 500.000 plants appear on the website, which received about 120.000 visits per month on average in 2021. We expect that appearing on the website (input) improves the visibility/reputation of plants (direct output), which may in turn increase the number of unsolicited applications received as well as the quality of applicants and/or the quality of the match between plants and applicants (intermediary outcomes). This may in turn have a positive impact on (i) the number of apprenticeships signed; (ii) the quality of these apprentices and (iii) the short and long term retention rates in the plants (final outcomes).[4] These effects will be measured both at the micro (plant) and macro (employment area) level (see below).

One potential drawback of the LBA website, however, is that the number of apprentices hired in a given plant because of the website may occur at the expense of other kinds of workers/work contracts. To explore the existence of such substitution effects, we will also, as a complementary analysis, look at the impact of the program on the evolution of standard unskilled employment in the treated and control plants, both at the micro and macro level.

Measurement

To measure the direct output of the program (i.e., plants' visibility/reputation), we will rely on LBA's internal information system that allows us to track, for every single plant in the treatment group (i) the number of times the plant appears in the result of searches; (ii) the number of times users have clicked on the information links for this plant ; (iii) the number of unsolicited applications directly sent to the plant *via* the website.

[3] A full description of the algorithm is available here :
https://mission-apprentissage.gitbook.io/la-bonne-alternance/dou-viennent-les-donnees

[4] Short-run retention rates are defined by the share of contracts broken within 6 months as well. Long-run retention rates are the share of apprentices remaining in their plant as a standard worker upon completion of their training. These outcomes inform us on the quality of the match between the apprentice and the plant as well as on the quality of the training provided by the plant. With these outcomes, we can check whether the site fosters a turnover of apprentices, to the detriment of high-quality apprenticeships (revealed by training-to-hire behaviours).

To measure intermediary outcomes, we plan to rely on a survey administered by phone and/or email to a sample of plants.[5] This would allow us to measure the extent to which appearing on the website leads plants to receive more unsolicited applications for apprenticeships, better applications and/or applications more adapted to the plant.

Finally, two main administrative datasets will be used to measure the final outcomes: the data on apprenticeship contracts (Ari@ne and DECA) as well as the social security data (DADS and DSN). These datasets will allow us to measure the number of apprentices hired in all establishments included in our study population. It will also allow us to measure the quality of these hires, which will be measured in two ways. The quality of the apprentice will be proxied by the level of the diploma s/he prepares. The quality of the match between the employer and the trainee is proxied by the probability that the contract is broken before its end. These datasets will also allow us to measure whether the apprentices keep on working in the same establishment after the end of their apprenticeship contract, as regular workers (either through a fixed-term contract or through an open-ended contract), which informs us on the quality of the match as well as on the quality of training.

We will first study the impact of the website on the intermediary and final outcomes at the micro (establishment) level. In addition, our two-step randomization design will also allow us to measure the effects of the program on these outcomes at the macro level (see Section 2). The same outcomes and measurement will be used for this macro-level analysis.

II. Evaluation Design

a. Study population

Our study population includes the exhaustive list of firms with a large potential of hiring on the apprenticeship market or on the standard labour market for at least one occupation, as identified in May 2021 by Pôle Emploi, the French Public Employment Service, by the means of the machine-learning algorithm previously described. This sample includes about 500,000 firms in total. Independently of this list, the website also highlights plants that have a vacancy opened for an apprenticeship contract. We will check whether the firms from this group are not disproportionately found among the control or the treatment group and will remove them from the analysis.

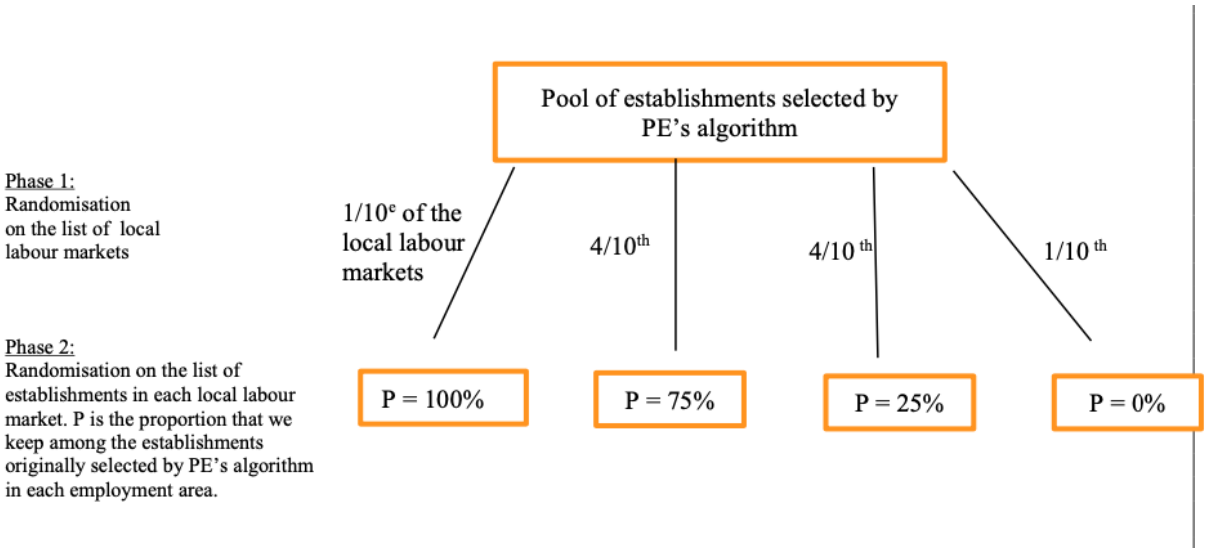
b. Randomization

The estimation strategy consists in inhibiting some of the establishments displayed on the website in a two-step randomisation design à la Crépon et al (2013). First, we randomly draw the proportion of plants to remain on the website (0%, 25%, 75%, 100%) in each of the 420 local labour markets (“bassins d’emplois”). Second, we randomly select plants to remain on the website according to the relevant proportion for their local labour market.

[5] As of today, discussions are still ongoing to validate the implementation of this survey.

The randomisation at the level of plants allows us to identify the effect of the website on the number of apprentices recruited and on the quality of the matches between employers and apprentices for a given establishment. The randomisation at the level of local labour markets allows us to study the "macro" effect of the website on the number and quality of apprenticeship contracts generated overall. This takes into account potential spillovers related to displacement effects. This second level allows one to know whether the (expected) greater number of apprentices recruited by the treated companies and/or the (expected) better quality of the "matches" occurs to the detriment of the competing plants from the control group in the same local labour market and whether the net effect of the website is to create contracts and/or improve their quality.

In practice, we randomly selected 1/10th of the local labour markets where to display 100% of the companies selected by PE's algorithm, 4/10th of the local labour markets where to display 75% of the establishments selected by the algorithm, 4/10th of the areas where to display 25% of the establishments selected by and the algorithm, and 1/10th of the areas where 0% should remain (see graph below).



III. Internal validity checks

a. Statistical methods

Estimation method

For each of the variables listed below, we will run an OLS regression of that variable on a dummy taking the value 1 for plants in the treatment group. To ensure that the coefficient of the dummy for treatment is the average of comparisons between treated and control firms within the same lottery, the regression will be weighted by propensity score weights. Specifically, let D denote the treatment group dummy, and let S denote the lottery in which a plant was included. Each treated plant will be weighted by $P(D = 1)/P(D = 1|S)$. Each control plant will be weighted by $P(D = 0)/P(D = 0|S)$. All regressions will also control for strata fixed effects. To test for the presence of externalities, we will follow Crépon et al (2013) and

estimate two models that are similar to their equations (6) and (7), respectively corresponding to an unconstrained and to a pooled reduced-form model.

Adjustment for multiple testing

For each variable/outcome, we will report both the unadjusted p-value of the coefficient of the treatment variable(s), and the p-value adjusted for control of the False Discovery Rate (see Benjamini & Hochberg, 1995) within each hypothesis. For each hypothesis containing several outcomes, we will also average those outcomes into a standardized measure, following the same standardization method as in Anderson (2008). This procedure consists in the following steps:

- For all outcomes, switch signs where necessary so that the positive direction always indicates a "better" outcome.
- Define J groupings of outcomes (also referred to as areas or domains). Each (normalized) outcome y_{jk} is assigned to one of these J areas, giving K_j outcomes in each area j, with k indexing outcomes within an area.
- Create a new variable, s_{ij}, that is a weighted average of y_{ijk} for individual i in area j. When constructing s_{ij}, weight its inputs—outcomes y_{ijk}—by the inverse of the covariance matrix of the transformed outcomes in area j.

We will then estimate the effect of the LBA website on that measure, the so-called standardized treatment effect.

Standard errors

Following Abadie *et al.* (2017), we will cluster standard errors at the level at which treatment assignment is determined, namely at the individual plant level for regressions estimating the micro (plant-level) effect and employment area level for regressions estimating the macro (employment area-level) effect.

b. <u>Hypothesis tested</u>

3.b.1 Differential attrition

H1a: The share of plants who are observed both before and after the end of the experiment is balanced between the treatment and the control groups

H1b: The share of plants who are observed both before and after the end of the experiment are balanced across the four treatment groups (0%, 25%, 75% and 100%).

3.b.2 Tests of covariate balance at baseline

H2a: At baseline, the average characteristics of employment areas are balanced across the four treatment groups (0%, 25%, 75% and 100%) :

1. Size of the area (in km2)
2. Population living in the area
3. Number of plants in the area included in the randomization
4. Share of the employment area's population living in a rural area
5. The unemployment rate in the area
6. Average number of apprenticeship offers per inhabitants over the 2019-2021 period
7. Average number of apprenticeship offers per unemployed over the 2019-2021 period
8. Average number of hires in apprenticeship contracts over the 2015-2021 period
9. Average number of hires in "contrats de professionnalisation" or apprenticeship over the 2015-2021 period[6]
10. Average number of hires in fixed-term or open-ended contracts over the 2015-2021 period

H2b: At baseline, the individual characteristics of plants are balanced between the treatment and the control group :

1. Number of hires in apprenticeship contracts over the 2015-2021 period
2. Number of hires in "contrats de professionalisation" or apprenticeship over the 2015-2021 period
3. Number of hires in fixed-term or open-ended contracts over the 2015-2021 period
4. Plant size
5. A dummy indicating location in a rural area
6. A dummy indicating availability of contact information on the LBA website
7. Number of jobs to which the plant is associated to on the LBA website

H2c: At baseline, the individual characteristics of plants are balanced across the four treatment groups (0%, 25%, 75% and 100%) : same variables as H2b.

IV. Effect of the program on the full sample

a. Statistical methods

Estimation method

The estimation method will be the same as in Section 3, except that the regressions will include control variables.

To select those control variables, we will use the method proposed in Belloni et al. (2014). We will select both plant-level controls and employment area-level controls. To select the

[6] "Contrats de professionnalisation" are another type of contracts that involve part-time studies at school and part-time work in a training plant..

plant-level controls, we will run a first Lasso regression of each outcome on all the plant-level variables listed in subsection 3.b.2, and a second Lasso regression of the treatment status of the plant on the same variables. The plant-level controls will be the variables selected either in the first or in the second Lasso regression. To select the employment area-level controls, we will run a first Lasso regression of employment area-average outcome on employment area-level variables listed in subsection 3.b.2, and a second Lasso regression of the treatment status of the employment area on the same variables. The employment area-level controls will be the variables selected either in the first or in the second Lasso regression.

In the Lasso regressions and in the final regressions where the effect of the treatment is estimated, missing values of the controls will be replaced by the mean of these controls, and for each control, a dummy for observations for which this control is missing will be included in the regression.

b. Hypothesis tested

4.b.1 Tests of plants' outcomes balance between the treatment and the control group at endline

H3a: At endline, the number of apprentices hired is balanced between treated and control plants

H3b: At endline, the quality of apprentices hired is balanced between treated and control plants

H3c: At endline, the retention rates of apprentices hired is balanced between treated and control plants

4.b.2 Tests of plants' outcomes balance between the four treatment groups at endline

H4a: At endline, the number of apprentices hired is balanced across the four treatment groups

H4b: At endline, the quality of apprentices hired is balanced across the four treatment groups

H4c: At endline, the retention rates of apprentices hired is balanced across the four treatment groups

V. Heterogeneity analysis

Estimation method

As a baseline estimation method for the heterogeneity analysis, we will use the same approach as in subsection 3.a. To complement this approach, we also plan to use machine learning techniques (e.g. Athey, Tibshirani & Wager (2019)), that will allow us to include

more dimensions in the heterogeneity analysis, and to focus on the 2-3 most important dimensions.

Standard errors
Same as in subsection 3.a.

Dimensions of heterogeneity analyzed
Apprenticeship training is not necessarily profitable for firms: training costs and the uncertainty regarding trainees' quality or the possibility to retain the apprentice upon graduation can exceed the discounted sum of present and future benefits (Dionisius et al, 2009). Firms' propensity to train reveals their expectations about the cost-benefit of hiring apprentices in their own context (Brébion, 2019a). We therefore expect the website to have a stronger effect on firms that are used to taking on apprentices. To test this hypothesis, we will estimate the heterogeneity of our main effects according to whether establishments have trained apprentices in the recent past. This will inform us about the most important margin on which the website operates: does it lead plants who already train apprentices to recruit and train more apprentices (intensive margin) and/or does it lead plants who usually do not train apprentices to do so (extensive margin)?

Among firms hiring apprentices, two main strategies can be distinguished according to whether firms seek profits *during* or *after* the apprenticeship period (Lindley, 1975; Wolter & Ryan, 2011). The first strategy consists in taking advantage of the low labour costs *during* the apprenticeship. The second strategy consists in training to hire: firms incur a positive net cost during the apprenticeship (to properly train the apprentice) and gather profits *afterwards*, by retaining the apprentice upon graduation at a time when s/he is fully productive. Given the large excess of labour supply on the apprenticeship market, one can expect that firms willing to follow the first strategy hire the number of apprentices they seek independently of the website. The LBA website should therefore have a stronger impact on firms training to hire: they are more selective in their hiring process and the increased visibility provided by the website should change their hiring choices if it leads to larger number/higher quality of unsolicited applications. We will therefore study whether the effect of the website depends on past retention rates of firms (see footnote 2 above for a definition of retention rates).

We also expect the effects of the website to be heterogeneous according to plant size. In particular, small firms have little resources to put into screening unsolicited applications and rather use local networks of their own workers or managers to hire, as suggested by de Larquier and Rieucau (2015). The website may have a stronger effect on large firms for this reason. Conversely, it could be the case that the impact of the website is stronger for smaller plants if there is a negative correlation between: (i) the ratio between the number of unsolicited applications firms receive thanks to the website and the usual number of unsolicited applications received and (ii) the size of firms. Likewise, the effect of the LBA website is likely to differ across economic sectors. In particular, a disproportionate share of apprentices at the secondary level is working in the craft industry whereas apprentices engaged in higher education are more often working in large firms of the service or manufacturing industry (Brébion 2019b). On the one hand, the literature has shown that, relative to responses to job postings, employers consider that unsolicited applications reveal a strong motivation of candidates (Bonoli & Hinrichs, 2012). Given the lower importance of diplomas in the hiring process of apprentices in the craft industry, one can expect the website to play a stronger role in this sector. On the other hand, candidates applying for an apprenticeship contract at the secondary-education level may lack the skills to put together a

strong unsolicited application to attract the interest of a manager. In such a case, the heterogeneity of our results would play the other way around, with a limited effect of the website in the craft sector.

In addition to this, the effects of the program are likely to be heterogeneous depending on the degree of market competition faced by plants. In particular, the potential negative spillovers are likely to be stronger in employment areas where the degree of competition is higher. To investigate this issue, we first plan to contrast the effect of the program for establishments located in rural areas vs in urban areas. Second, we will explore whether the effects differ depending on a measure of market competition (e.g., Herfindahl-Hirschman Index). Third, we will scrutinize the role of youth unemployment (measured at the local level) to explain the strength of the effects.

Finally, the apprenticeship market is very diverse and we expect that firms receiving the most relevant applications will better value the flow of applications stemming from the website. We may gain access to some data about the demographics of the candidates and the content of the cover letters sent through the website, which would allow us to test our hypothesis. Thus, this data will be used to describe the average candidate using the website (education level and age in particular), which was uncertain at the start of the RCT. This will allow us to test whether the effect of the website for a given firm increases with the quality of the fit between the characteristics of this average apprentice and those of the apprentices usually hired by that firm.

References

Abadie, A., Athey, S., Imbens, G.W., et al. (2017) When Should You Adjust Standard Errors for Clustering? 24003. *National Bureau of Economic Research*. doi:10.3386/w24003.

Anderson, M. L. (2008), 'Multiple inference and gender differences in the effects of early intervention: A reevaluation of the abecedarian, perry preschool, and early training projects', *Journal of the American statistical Association,* 103(484), 1481–1495.

Athey, S., Tibshirani, J., & Wager, S. 2019. "Generalized random forests". *The Annals of Statistics*, 47 (2):1148–1178

Belloni, A., Chernozhukov, V. and Hansen, C. (2014) High-Dimensional Methods and Inference on Structural and Treatment Effects. *Journal of Economic Perspectives*, 28 (2): 29–50. doi:10.1257/jep.28.2.29.

Benjamini, Y. and Hochberg, Y. (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society*. Series B (Methodological), 57 (1): 289–300

Bonoli, G., & Hinrichs, K. (2012). Statistical Discrimination and Employers' Recruitment. *European Societies*, *14*(3), 338–361.

Brébion, C. (2019a). *Vocational training and industrial relations in France and Germany* [PhD Thesis].

Brébion, C. (2019b). L'apprentissage, un meilleur « rendement » professionnel en France qu'en Allemagne. *Formation emploi. Revue française de sciences sociales*, *146*, 101–127.

Crépon, B., Duflo, E., Gurgand, M., Rathelot, R., & Zamora, P. (2013). Do labor market policies have displacement effects? Evidence from a clustered randomized experiment. *The Quarterly Journal of Economics*, *128*(2), 531-580.

Dionisius R., Muehlemann S., Pfeifer H., Walden G., Wenzelmann F. & Wolter S. C. (2009), "Costs and benefits of apprenticeship training. A comparison of Germany and Switzerland", *Applied Economics Quarterly*, 55(1), pp. 7-37.

de Larquier, G., & Rieucau, G. (2015). Candidatures spontanées, réseaux et intermédiaires publics: Quelle information et quels appariements sur le marché du travail français ? *Relations industrielles / Industrial Relations*, *70*(3), 486–509.

Lindley, R. M. (1975). The Demand for Apprentice Recruits by the Engineering Industry, 1951-71. *Scottish Journal of Political Economy*, *22*(1), 1–24.

Wolter, S. C., & Ryan, P. (2011). Chapter 11—Apprenticeship. In E. A. Hanushek, S. Machin, & L. Woessmann (Eds.), *Handbook of the Economics of Education* (Vol. 3, pp. 521–576). Elsevier.