Online Platforms and the Labour Market: Learning (with Machines) from an Experiment in France

Master Thesis

Yagan Hazard

Supervisor: **Luc Behaghel**

Referee: **Maxime Tô**

Master Public Policy and Development

Paris School of Economics

June 2019

Abstract

I study the effect of an online job search assistance program taking advantage of a previous experiment made by the French public employment services, which provides some exogenous variation in the use of this platform. I focus on the heterogeneity analysis of this treatment, using two main different approaches.

The first one is theory-driven, and focus on the analysis of the heterogeneity of the treatment with respect to various different labour market tightness indicators. Two main assessments can be made based on this analysis. (i) Tightness indicators are (surprisingly) decorrelated, making it difficult to corroborate the rare significant results obtained. (ii) The set of significant results obtained suggest that the treatment effect is *increasing* in labour market tightness. I suggest competing ways of modelling the treatment consistent with those results. I also document some evidence of a larger treatment effect for individuals with weaker employment prospects. This is in line with other empirical evidence in the literature evaluating job search assistance programs.

The second approach is more data-driven, and resorts to the new machine learning (ML) techniques developed for heterogeneity analysis. I focus on tree-based techniques and forests, which have been central in the development of these techniques. The results of this analysis shed light on the limits of ML in the exploration of treatment effect heterogeneity, especially as the main ML-specific test for treatment effect heterogeneity developed by Chernozhukov et al. (2018a) concludes that ML is unable to detect any heterogeneity — yet this might be not that surprising after all given the lack of statistical power (low take-up) and the probably low order of magnitude of the treatment effect studied. Still, I provide applications of a large part of the existing ML techniques for treatment effect heterogeneity, trying to take advantage of each of them to document which are the dimensions that are likely to be important to study treatment effect heterogeneity in my setting.

Keywords: online job search, mismatch, market tightness, vacancies, machine learning
JEL Codes: E24, J62, J64, C18

Remerciements

En premier lieu, je suis profondément reconnaissant à Luc Behaghel pour avoir accepté de diriger ce travail, pour les discussions passionnantes que cela a pu susciter, et pour sa grande disponibilité. Je tiens également à remercier Maxime Tô pour avoir accepté d'être le rapporteur de ce mémoire; et l'ensemble des professeurs de l'Ecole d'Economie de Paris ayant accepté de m'aider en dialoguant à ce sujet. Je remercie en outre grandement Pôle emploi pour m'avoir autorisé à utiliser les données de ce mémoire, et Juliet Uraz pour l'extraction desdites données et pour de précieux conseils quant à leur utilisation.

Ensuite, j'ai une pensée pour les nombreux camarades avec qui j'ai pu travailler tout au long de ce master ; ces deux années furent éprouvantes, et j'ai conscience de tout ce que je dois au soutien que vous m'avez apporté. Merci plus particulièrement à Simon, sans qui ce master n'aurait pas été le même — et par là, j'entends qu'il aurait sans doute pu se transformer en calvaire par moment ; merci pour tous ces moments incroyables à se retourner ensemble le cerveau avec l'obsession que l'on nous connaît ; merci pour m'avoir permis de me dépasser bien au-delà de ce qu'il m'aurait été permis en ton absence. Je pense également à Thomas, qui ne s'est jamais lassé de nos discussions concernant ce mémoire, et à toujours su trouver les mots (ou les références judicieuses) pour me relancer dans ce travail.

Enfin, *last but not least*, je veux remercier mes proches, qui m'ont supporté si longtemps dans mes études et plus particulièrement au cours de ce master malgré les doutes, les coups de stress, et autres péripéties estudiantines. La taille de ce paragraphe n'est pas à la mesure de la gratitude que j'ai pour vous.

Contents

1 Introduction

The aim of this thesis is to learn about an active labour market policy, namely online job search assistance. I take advantage of a previous experiment made by the French public employment services (Pôle emploi, PE henceforth) in 2015. The experiment consisted in an evaluation of the online platform named *La Bonne Boîte* (LBB henceforth): around 75 000 job seekers randomly selected were invited by e-mail to use the LBB online platform to search for jobs, while around 75 000 other job seekers were randomly selected as control units. This experiment is a rather rare opportunity to learn about online job search, as there is not so much experimental evidence of the effect of online platforms on job search (see Belot et al. (2018b) and Horton (2017) for the two examples I know about). This may be surprising as (i) experimenting online job search assistance is arguably less costly than experimenting intensive counselling, which has been done many times (see Behaghel et al. (2014) for a review of this literature), and (ii) there has been some hope for a long time that the rise of the Internet would have an important impact on job search, allowing for a reduction of frictions because of the massive information diffusion allowed (see for instance Autor (2001) for an early sketch of these hopes). One candidate explanation for this is the disappointing results of early studies on the subject. Indeed, in one of the first papers evaluating the effect of online job search (in a non-experimental setting, using observational data from CPS), Kuhn and Skuterud (2004) found that individuals reporting to use the Internet for online job search were *less* likely to find a job, even when controlling for a detailed set of worker and labor characteristics, along with an individual overall search intensity measure. Some explanations offered by the authors were the following: (i) the low cost of online job search could attract individuals with lower unobserved motivation and/or (ii) individuals using online job search are using it because of a lack of other job search channels that are more effective than online job search, such as informal contacts and social network — this follows Granovetter's "strength of weak ties" theory. This results of ineffectiveness of online job search was even confirmed by another study later in the 2000s, as Kroft and Pope (2014) used the expansion of an online job search platform (Craigslist) in various US cities as a quasi-experimental setting, and found no effect on local unemployment rates. It was only recently that hopes were revived, notably after Kuhn (2014) work that found, using a similar methodology as in Kuhn and Skuterud (2004), the reversed result: individuals using the internet were found *more* likely to find a job than others. Candidate explanations for this are (i) that online platforms improved greatly their product over a decade, allowing for a more efficient search (less fake/stale résumés for instance), and (ii) probably even more importantly the network externalities resulting from the wider use of the Internet: more individuals using this network may have benefited to all users, as the Internet is useful to job seekers if a large enough number of vacancies are posted on it, and to firms if a large enough number of job seekers consult their ads and post their résumés on it. After this article came the first experimental evidence about job search. Firstly, Horton (2017) shows, using random recommendations of candidates to firms, that giving such advice to firms was increasing significantly their vacancy fill rate, without evidence of crowding out of non-recommended applicants. Additionally, the author finds that this effect is larger for vacancies likely to receive a smaller pool of applicants, which seems in line with basic insights of the equilibrium unemployment theory (see below). A second piece of experimental evidence on this topic is provided by Belot et al. (2018b). Developing their own job search platform, they invited a sample of job seekers to use it in order to evaluate whether giving them some tailored information (randomly given to part of the job seekers) about available job vacancies could affect their research behaviour. They find that customized advice leads to a broadening of the set of jobs considered, resulting in a larger number of job interviews.

The theoretical background that has motivated this whole empirical literature is the one of matching models from the equilibrium unemployment theory (see Rogerson et al. (2006) and Petrongolo and Pissarides (2001) for extensive surveys of this literature, and Pissarides (2000) for a comprehensive exposition of the equilibrium unemployment theory). Indeed, all expectations about the effect of the Internet of the labour market and job search are linked to the idea of a frictional labour market, in which part of the unemployment is due to an inefficient matching between workers and firms, that could potentially be improved by online platforms if they ease meetings between agents. A key quantity involved in these models is labour market tightness, i.e. the ratio between the number of vacancies and the number of unemployed on the labour market. In these models, this quantity enters in a rather simple and intuitive way in the determination of the matching rate between vacancies and job seekers, the job finding probability of job seekers, and the equilibrium unemployment level. Therefore, it is tempting to study the effect of job search assistance through the lens of this theoretical framework. In particular, it might be interesting to study the interaction existing between the effectiveness of online job search assistance and labour market tightness. However, this approach is not as simple to implement as it may seem, as it raises (i) measurement issues — regarding labour market tightness indicator(s) — and (ii) identification and inference issues — how to determine the right model to test, and to distinguish between false discovery induced by multiple hypothesis testing and causal relationships ? Indeed, there are several alternatives to measure labour market tightness in France (see for instance Orand (2018) for a study of these different measures), that are highly uncorrelated, indicating that these indicators do not capture the same information about labour markets. How to choose between one or another when interacting labour market tightness with the treatment in a classical approach to study heterogeneous treatment effect ? If testing for each of these indicators, aren't we exposed to the issue of data snooping and multiple hypothesis testing ? How to choose the control variables that should be included when studying this heterogeneity in a classical OLS ?

Faced with these challenges, machine learning (ML henceforth) and associated techniques (data splitting, cross-fitting etc.) appear more and more as candidate solutions. There is a growing number of empirical papers resorting to ML methods in order to study heterogeneous treatment effects. In particular, Knaus et al. (2018) and Strittmatter (2019) use these new tools to study heterogeneity in active labour market treatment effects — Knaus et al. (2018) focusing especially on job search programmes. In this way, I will attempt to contribute to this literature by implementing various heterogeneity exploration techniques, from standard ones — using interaction terms in OLS — to the newest ones — such as causal trees (Athey and Imbens, 2016a), causal forests (Wager and Athey, 2017), and Best Linear Predictor (BLP), sorted Group Average Treatment Effects (GATEs), CLassification ANalysis (CLAN) developed by Chernozhukov et al. (2018b).

The rest of this thesis is organized as follows. The next section provides a literature review of the different recent ML techniques developed in order to uncover treatment effect heterogeneity. Section 3 presents the experiment and describes the data, along with some descriptive statistics. Section 4 presents a model of job search, in order to propose a candidate framework to read some of the results. Section 5 presents first results analysing the treatment and exploring in classic ways treatment effect heterogeneity, and Section 6 presents the results from implementing new ML techniques to explore heterogeneity. Section 7 concludes.

2 Related Literature

> "There are two cultures in the use of statistical modelling to reach conclusions from data. One assumes that the data are generated by a given stochastic data model. The other uses algorithmic models and treats the data mechanism as unknown. The statistical community has been committed to the almost exclusive use of data models. This commitment has led to irrelevant theory, questionable conclusions, and has kept statisticians from working on a large range of interesting current problems. Algorithmic modelling, both in theory and practice, has developed rapidly in fields outside statistics. It can be used both on large complex data sets and as a more accurate and informative alternative to data modelling on smaller data sets. If our goal as a field is to use data to solve problems, then we need to move away from exclusive dependence on data models and adopt a more diverse set of tools."
>
> — Breiman (2001), p.199

In the recent period, machine learning methods have become more and more popular in the applied economics field (see Mullainathan and Spiess (2017), Bertrand et al. (2017) or Strittmatter (2019) for instance). The above quotation, an abstract from an article by Leo Breiman at the beginning of this century, summarises rather well what is at play here[1]. Given the fast development of computer power and the opportunities it opens for ML ("algorithmic models" in Breiman's words), along with the access to larger and larger data sets, the statistical community cannot ignore any longer these data-driven approaches and has to adapt these tools in order to apply them to their own goal, i.e. "[using] data to solve problems". Varian (2014), and Mullainathan and Spiess (2017) after him, provided some first influential assessments of the usefulness of "off-the-shelf" ML techniques for econometrics and causal inference: pre-processing of complex data sets (such as satellite images, text etc.), or prediction steps in parameter estimation (flexible prediction of propensity scores for matching estimators for example), or policy targeting (predicting heterogeneous treatment effects of a program in order to offer a program to individuals with the highest predicted impacts) among other things. ML is perfectly suited for these tasks since they take advantage of its main strength, namely out-of-sample prediction. However, econometric problems typically revolve around identifying causal relationships, which is not *a priori* what ML techniques were developed for in the first place. Indeed, causal inference

[1] This quotation has also been used in the introduction of Athey and Imbens (2019).

is usually linked to data modelling: under some assumptions made on the data generating process, one can identify and estimate a parameter that describes the causal relationship between two variables. Therefore, a simple question arise: how to benefit from the predictive power of ML techniques, along with their flexibility and their ability to deal with high-dimensional data sets, while still aiming at uncovering causal links? Athey and Imbens (2019) provide a very clear and rather exhaustive review of all recent ML techniques developed with this goal of filling (progressively) the gap between those "two cultures".

In this regard, a first step has been the development of ML tools in order to identify some key (treatment effect) parameter. An important paper is the one by Belloni et al. (2014) about double post-selection LASSO. The idea is rather simple: using ML techniques, and in particular Least Absolute Shrinkage and Selection Operator (LASSO) combined with data splitting, in order to implement an optimal selection among numerous covariates so that in the end, a key parameter (such as the effect of a program) could be identified cleaned of omitted variable bias. This technique has been extended to the use of other ML techniques than LASSO, in Chernozhukov et al. (2018a). Their "double/debiased machine learning" (DML henceforth) proposes to use any ML method (random forests, LASSO, ridge, deep neural nets, boosted trees, ensemble methods etc.) to predict flexibly conditional means of both the outcome and the variable of interest, on a given split of the data. Then these predictions are fitted on the other split of the data, and one can regress the residual of the outcome on the residual of the variable of interest in order to get a estimate of the causal effect of this variable on the dependent variable. Ultimately, in order to reach efficiency, the role of each split of the data is reversed (this is what is usually called "cross-fitting"), and the ultimate estimate is given by the mean of the estimates obtained on each split. As well putted by Duflo (2018), there is "no magic here": this is "Frisch-Waugh for the 21st century, [and] the causal estimate is only as good as the covariates". Indeed, this is only a way to deal flexibly with a large number of covariates in a regression: if all potential confounders are observed, then DML should manage to find the right way to control for them — while OLS could not deal with too many covariates, therefore limiting the flexibility of the conditional expectation function estimated. Another alternative in the "classic" econometric literature is matching estimators, yet as Athey and Imbens (2019) underscores, "their theoretical and practical properties are poor with many covariates". That said, ML does not allow for miracles: if there are important unobserved confounders, obviously DML will not be able to get rid of omitted variable bias.

Another important task at which ML can arguably be put to good use is uncovering treatment effect heterogeneity. A series of recent papers have been focused on this theme in the past few years, and this is seen as a prime example of how ML can be thoughtfully adapted to answer to econometric issues. Determining treatment effect heterogeneity is often of first importance in policy evaluation: it allows to determine under which conditions a program will be the most (or the least) efficient, so that an optimal assignment rule could be derived in order to maximise the benefits derived from the treatment for a given budget (see Athey and Imbens (2017) for an argument on this subject). However, when there are many covariates, meaning many different possible splitting of the data, researchers face the threat (or temptation?) of data mining: they might end up focusing on a specific split of their original sample on which they find a large/significantly larger treatment effect. Then the issue is that part of the heterogeneity uncovered is likely to be an artefact that holds only for the specific draw at hand, and valid inference is no longer possible. A way to deal with this "multiple hypothesis testing" (MHT henceforth) issue is an exhaustive search for heterogeneous treatment effects, followed by a correction for the risk of "false discovery" described above. There is a rather wide literature on the subject (see Farcomeni (2007) for a review), yet these methods are not that often implemented in empirical economics. The reason is rather obvious: classic MHT correction methods tend to be quite conservative, and as a consequence applying them in social sciences often ends up failing to reject (nearly) all null hypotheses — which makes those techniques arguably not that helpful to enhance scientific discovery. A recent work by List et al. (2016) proposes a new MHT correction method adapted to experimental economics, with the goal to make it less conservative. However, this technique still requires for the researcher to specify in advance all the hypotheses he/she is willing to test, which might be a restrictive way to look for heterogeneity since the researcher is likely to limit the flexibility of his/her tests: what interactions should be tested? How many polynomials should be included in the set of covariates? Athey and Imbens (2017) offer a useful description of the intuition behind List et al. (2016) method, and express those concerns. At this point, the question is: what can ML do in this regard?

Athey and Imbens (2016a) proposed to adapt a classic ML procedure, namely the regression tree (Breiman et al., 1984), in order to make it look for partition of the data that would present heterogeneous treatment effects — the resulting algorithm was coined by the authors as a "causal tree". In order to do that, they change the decision criterion used in order to build the tree: instead of focusing on mean-squared error of outcome prediction, their tree is built focusing on mean-squared error of treatment effects. Moreover, the algorithm is such that only half of the data is used to build the tree, while the other half is kept for estimation of treatment effect and confidence intervals for each "leaf" (i.e. each subgroup created by the tree): this prevent from any concern regarding overfitting when estimating conditional average treatment effects (CATEs henceforth) within each leaf, hence it allows for valid inference while investigating very flexibly, with the help of the machine, treatment effect heterogeneity. In addition, the output of this procedure is very easy to interprete — Athey and Imbens (2017) compare it to "a decision tree [a doctor would use] to determine whether to prescribe a drug to a patient". However, there are several drawbacks to this approach. First of all, experience has shown "inconsistent model selection": depending on the split of the data considered, the tree built by the ML algorithm is different, and therefore it is hard to determine the relevant procedure to avoid digging into different data splits up until one finds his/her preferred tree to present. Second, there are still some tuning parameters to choose to build a tree: the authors are now working on some algorithms in order to tune automatically each parameters based on cross-validation[2], but these are not fully functioning for now (and are quite computationally costly[3]). Lastly, this procedure does not allow for individual treatment effect estimation: one could wish a CATE function that would be more flexible[4]. Wager and Athey (2017) offered a way to reach such a function using random forests — leading to the creation of a "causal forest" algorithm, which amounts to the creation of multiple "causal trees" which treatment effect predictions are averaged in order to create a CATE function that varies smoothly with covariates, and gives an individual treatment effect estimate (if enough trees are grown)[5]. In addition, the authors show that their algorithm allows for the estimation of confidence intervals for each treatment effect estimate. Yet the causal forest option has the defaults that go with its qualities: it allows for a more flexible, "smooth" and individualized CATE function, but this very function becomes much more obscure than its causal tree version. How to dig into this black box to get a sense of what is driving treatment effect heterogeneity, if it exists? In her lectures posted on the American Economic Association (AEA) website[6], Susan Athey proposes to create some heatmaps and/or to study covariates values along CATE quantiles to grasp some key features of the CATE function. But a limit to this approach is that it exposes again to MHT criticisms, since at this stage the researcher has

[2]See `tune.parameters` option in `causalTree` R package.
[3]This is even more true for forests than for trees.
[4]To be clear, by CATE function I mean a function such as $\tau(x) = E\left[\tau_i | X_i = x\right]$, where $\tau_i = Y_i(1) - Y_i(0)$
[5]See `grf` R package for implementing the causal forest algorithm.
[6]See: `https://www.aeaweb.org/conference/cont-ed/2018-webcasts`

to choose by himself the variables used to dig into the CATE function, hence there is no longer any valid inference strategy at this point. A late development in the study of the CATE function is the elaboration of some "variable importance" measures — see O'Neill and Weeks (2018) for a brief description of the different variable importance measures that exists along with their pros and cons. A rather intuitive one amounts to a count of the proportion of splits made using the variable at hand up to a depth of 4, with a depth-specific weighting (splits made at a lower depth in a tree have a larger weight). The formula for the importance of covariate X_j is then given by[7]:

$$\text{importance}\left(X_j\right) = \frac{\sum_{k=1}^{4}\left[\frac{\sum_{\text{all trees}} \text{Nb. depth-k splits on } X_j}{\sum_{\text{all trees}} \text{Tot. nb. depth-k splits} k \text{ splits}}\right] k^{-2}}{\sum_{k=1}^{4} k^{-2}}$$

I am not aware of serious discussions of the properties of the different existing variable importance measures.

Another approach to the CATE function investigation problem has been developed by Chernozhukov et al. (2018b), with the idea to aim at "simple" features of this function. They develop valid inference for three main objects[8], namely (using the authors' notations):

i. the Best Linear Predictor (BLP) of the true CATE function $s_0(Z)$ based on the ML proxy predictor $S(Z)$;

ii. the sorted Group Average Treatment Effects (GATEs), the average of conditional treatment effects $s_0(Z)$ by heterogeneity groups (quintiles of predicted CATEs) induced by the ML proxy predictor $S(Z)$;

iii. the CLassification ANalysis (CLAN) that reports the average characteristics of the most and least affected units — as defined by the ML proxy predictor $S(Z)$.

This approach has its advantages. Firstly, it offers a valid test for the existence of any treatment effect heterogeneity. Indeed, taking the BLP identification regression formula:

$$Y = \alpha' X_1 + \beta_1(D - p(Z)) + \beta_2(D - p(Z))(S - \mathrm{E}S) + \epsilon$$

one can test whether there is treatment effect heterogeneity by testing whether β_2 is statistically significant from 0. If this is the case, then one can argue that there is indeed some heterogeneous

[7]The maximum depth considered (4 here) and the decay factor for differential weighting between different depth splits (the factor on k, 2 in the formula presented) can be modified: the values reported here are the default ones in the `variable_importance` function of `grf` R package.

[8]One could discuss the validity of the inference for CLAN, as the MHT issue is once again a potential concern at this stage.

treatment effects[9]. Similarly, one can test whether the ML proxy predictor $S(Z)$ provides some relevant heterogeneity groups by creating quantiles (quintiles for instance, as in Chernozhukov et al. (2018b) application) of predicted CATEs and estimating the ATE among each of these quantiles: this is the GATEs. Lastly, if one finds out that there is indeed a significant difference in ATE across heterogeneity groups as defined by the ML proxy predictor, one can study the average characteristics of the most and least affected heterogeneity groups, in order to get a sense of the covariates driving the heterogeneity. However, at this stage, one should again be worried about MHT when testing for significant differences in mean covariates values between groups.

There is a growing number of studies resorting to ML in order to study heterogeneous treatment effects — see Bertrand et al. (2017) in development economics, O'Neill and Weeks (2018) studying the effect of different electricity pricing schemes, or Knaus et al. (2018) and Strittmatter (2019) about active labor market policies. At this stage, it seems to me that despite this enthusiasm for these new tools, researchers are still searching for a conventional way to use them: most of these work still investigate the comparative advantages of each of these techniques, by implementing several of them. As a consequence, I will attempt in this work to contribute to this literature by presenting several different ML methods to investigate treatment effect heterogeneity in my setting.

3 Experimental Design and Data

3.1 The LBB platform and the experiment

The goal of LBB is to detect firms that are more likely to hire, independently from whether a firm has posted a vacancy at Pôle emploi or not. The rationale for this is that according to LBB, around 75% of hirings in France are made without any vacancy being posted at Pôle emploi:[10]; as a consequence, there would be a "hidden labor market", accessible only through spontaneous applications. And since it is difficult to determine to which firm such applications should be sent, the website offers to select the most relevant ones for job seekers, with the following headline: "Do not send applications at random anymore"[11]. In order to do that, LBB uses an admin-

[9]Note that on the contrary, failing to reject the null hypothesis $\beta_2 = 0$ does not allow to conclude that there is no treatment effect heterogeneity. It can simply be that the ML proxy predictor $S(Z)$ is useless in uncovering it.

[10]See F.A.Q. on LBB website (`https://labonneboite.pole-emploi.fr/faq` visited on 29/05/2019).

[11]In French: "*N'envoyez plus vos candidatures au hasard*".

istrative database recording all hirings made by french firms[12], and builds a statistical model predicting future hirings based on previous hirings observed in the administrative database[13]. These predictions are made for more than 500 different occupations[14]; based on these results, LBB reports on its website firms that have the highest chances to recruit within the next 6 months according to their model, for each of these 500 occupations. A job seeker can then fill in the occupation he/she is looking for and the geographic area of search, and LBB provides a list of firms corresponding to the search criteria[15].

The experiment consisted in an evaluation of the online platform named La Bonne Boîte (LBB henceforth). In October 2015, 200,000 job seekers were randomly selected among the pool of job seekers registered in "catégorie A" at PE[16]: 100,000 were supposed to be the control units, and the other 100,000 were supposed to be the treated units that would receive a e-mail from PE with a link directing towards LBB website, in order to encourage them to use it in their job search. However, the e-mails would only be sent on December 15$^{\text{th}}$; meanwhile, in both groups, some individuals would get out of "catégorie A" — for various reasons: job finding, entering in a training etc. As a consequence, experimenters had to exclude those individuals from both groups in order to keep two comparable groups: this leaves us with 154,779 individuals, with 77,282 control units and 77,497 treatment units[17].

Overall, this experiment follows a classic encouragement design: the e-mail is an instrument to encourage job seekers to use the LBB website. However, one should note that it is difficult to analyse the effect of the treatment considered beyond the Intention-To-Treat (ITT henceforth) effect. Indeed, even though I have some information about e-mail opening and whether a treated job seeker follows the LBB link inserted in the e-mail, I do not observe several important information in order to go beyond ITT analysis. First, it might be considered as a bold assumption to assume that individuals who did not clicked directly on the LBB link within the e-mail did not consulted the website at all; if this assumption does not hold, then the estimation of the

[12] This is the database of DPAE, "Déclaration préalable à l'embauche".

[13] I do not have access to the exact model used by LBB. Nonetheless, I know that this model is using a simple OLS in order to build their predictions. The model is not subject to a very strong validation for now (only a check that the root mean squared error is "not too large").

[14] To be precise, this includes all of the 532 ROME ("Répertoire opérationnel des métiers et des emplois") codes.

[15] If there are not any firms that match the search criteria, then LBB offers to provide a list of firms recruiting in an occupation close to the one indicated by the job seeker.

[16] "Catégorie A" is the category of job seekers in Pôle emploi that is the closest to the BIT definition of unemployment. In french, Pôle emploi defines this category as follows: "Personne sans emploi, tenue d'accomplir des actes positifs de recherche d'emploi, à la recherche d'un emploi quel que soit le type de contrat (CDI,CDD, à temps plein, à temps partiel, temporaire ou saisonnier)" (from Pôle emploi website: `https://www.service-public.fr/particuliers/vosdroits/F13240` visited on 03/06/2019).

[17] To be precise, once we exclude observations for which we have NAs for some variables used in the analysis, we are left with 149,286 observations: 74547 control units and 74739 treated units.

effect of using LBB (as proxied by clicking on the link) would be biased, despite the experimental design. Secondly, I do not observe whether control units use LBB in their search — meaning, in the usual terminology, that I do not observe whether there are "always-takers" in the control group. This would further bias the treatment effect analysis. For these reasons, I will focus in this work on ITT analysis — without loosing so much, since ITT stays the policy-relevant quantity to analyse. Furthermore, this choice is also motivated by the fact that ML techniques to explore heterogeneous treatment effects are not all adapted (for now) to instrumental settings — see Athey et al. (2018) for an example of adaptation of one of these methods (causal forests) to instrumental settings. That said, ITT and Treatment will be used equivalently in the rest of this paper — unless explicitly stated otherwise.

One should also have in mind that treated job seekers received the e-mail right before Christmas — on December 15$^{\text{th}}$ — and did not receive any extra e-mail as reminders. Moreover, the information in the e-mail did not contain any tailored information — such as personalised advice based on the kind of job the individual is looking for. These are candidate explanations for the very low "take-up" (see Table 1). Indeed, only 35% of treated individuals opened the e-mail, and only 19% (of the whole sample of treated individuals) actually clicked on the link. That might explains why there are difficulties to detect an average treatment effect based on this experiment (see Section 5.1), and motivates some extra analysis of treatment effect heterogeneity in order to establish (at least) if the treatment has significant effects on some particular groups and/or in some particular situations.

3.2 Data description

The primary data source is the administrative database from PE, the *Fichier Historique* (FH henceforth). I complemented it with different measures of labor market tightness, from different sources (DARES-PE, *Besoins en Main-d'Œuvre* (BMO) survey).

3.2.1 Data extracted from PE *Fichier Historique*

As evoked above, the main source of information come from PE administrative records (FH). These includes sociodemographic variables on all job seekers involved in the experiment (see Table 1 for a list of these variables). They also allow to construct the outcome variable(s) for this analysis, as they follow job seekers up to 24 months after the random assignment and e-mail sending: we therefore know, for each of these 24 months, whether any job seeker in the sample has find a job between the beginning of the experiment and this month. These outcome variables indicating "return to work"[18] are built by PE based on several sources. The final indicator that we have for each month is coming from PE database ICT 01[19]. The main outcome that I will consider in this study is whether there is a transition from unemployment to employment — as indicated by the IRE indicator — during the 6 months following the experiment.

Along with this file is merged the assignment file, that gives three information: (i) whether the individual is treated (i.e. e-mail is sent) or not; (ii) whether he/she opened the e-mail; (iii) whether he/she clicked on the link included in the e-mail, pointing to LBB website. For the reasons evoked by the end of Section 3.1, I do not use so much information (i) and (ii) and focus on ITT analysis.

3.2.2 Tightness data

As explained in Section 1, motivated by the unemployment equilibrium theory, I decided to study in particular the interactions between the treatment effect and labor market tightness. In order to do that I needed to enrich PE administrative records with some local labor market tightness data. Yet, there are no common indicator for this quantity. There are several reasons for this. First, while the denominator, the stock of job seekers on a market, can be obtained using PE administrative records, the numerator is way more complicated to measure: for instance, vacancies posted at PE do not represent all vacancies, and are likely to be a selected sample. Second, there is no common definition of local labor markets, both for the geographical and occupational delimitations: should labor markets be at the level of commuting zones[20], departments, regions? What should be the occupational classification used, and up to which granularity? All these questions do not have standard answers. Therefore, using the different data sources I could access in the context of this work, I tried to come up with the best measures of local labor market tightness. I mainly had to data sources: (i) local labor market tightness data produced by PE based on the *Statistiques mensuelles du marché du travail* (STMT) database, and used by the *Direction de l'Animation de la recherche, des Études et des Statistiques* (DARES) in its

[18] *Indice de Retour à l'Emploi* (IRE) in French.

[19] A note about the elaboration of the IRE in PE database ICT 01 (in French): L'indice de retour à l'emploi (IRE) calculé mensuellement, mesure le nombre de reprises d'emploi durant le mois m parmi les demandeurs en catégorie A et B au cours du mois $m-1$. Il est disponible dans le fichier historique archivé. Sont comptés comme une reprise d'emploi au sens de l'ICT01, le fait d'avoir une DPAE (déclaration préalable à l'embauche) de un mois ou plus, de sortir des listes pour une reprise d'emploi déclarée, de basculer en catégorie C ou E - sans retour en catégorie A ou B au mois suivant, ainsi que d'entrer dans une formation de préparation à la prise de poste - AFPR (action de formation préalable au recrutement) et POE individuelle (préparation opérationnelle à l'emploi).

[20] The equivalent would be INSEE *zones d'emploi* or PE *bassins d'emploi*.

publications[21], and (ii) the 2016 *Besoins en Main-d'Œuvre* PE survey.

Local labor market tightness data from PE STMT database This database, used by the DARES in its publications, records the flows and stocks of job seekers by regions — 3 different regions in my case, in the context of the experiment[22] — and by occupational code — using the *nomenclature des familles professionnelles* (FAP 2009) at its finer (5 digits) level, meaning 208 levels. The FAP classification is useful since it was explicitly built in order to be compatible with multiple other (french) occupational classifications. Among other things, it allows me to match it with the occupational classification used in PE records, which is the *Répertoire opérationnel des métiers et des emplois* (ROME)[23], and I was also able to use it for the other set of tightness indicators based on the BMO survey, as it is the classification used in this survey[24]. I chose to use the FAP classification at the 3 digits level for my analysis — which leaves me with 87 different codes. The only rationale for this is the following: my second set of indicators based on BMO allow me to go at a finer level of geographical disaggregation — departments (17 different departments in my experiment) instead of regions — so that with 208 different occupational levels, I would have had only few observations (or none) per market for these indicators. Since I wanted to keep the same occupational classification for all different tightness indicators used, I made this choice. Notice that there are with no doubts room for improvement in this regard: the selection of the different occupational codes could have been more data-driven — based on the dispersion of tightness indicators within/between occupational groups — or based on some qualitative information on the tasks used in each occupations; I did not have enough time to look into this (rather difficult and untreated) question. That said, all the tightness indicators based on PE STMT database are therefore defined at the region (3 levels) $\times$ aggregated FAP code (87 levels). I present below a list of the different indicators built based on this data source.

List of the different indicators based on PE STMT database with description:

[21] I am grateful to Sara Signorelli for accepting to share this data with me.

[22] *Lorraine*, *Pays-de-la-Loire* and *Île-de-France*.

[23] For a few observations (around 1,000 observations), there were multiple FAP codes possible for the ROME code of the observations. For these individuals, I had to attribute a FAP code randomly among the possible FAP codes.

[24] For a description of the FAP classification and crosswalks between FAP and other french occupational classifications (ROME, PCS), see: `https://dares.travail-emploi.gouv.fr/dares-etudes-et-statistiques/statistiques-de-a-a-z/article/la-nomenclature-des-familles-professionnelles-fap-2009`

- `tension_stock_PE`: the ratio between the stock of vacancies posted at Pôle emploi at the end of december 2015 and the stock of "catégorie A" job seekers registered at Pôle emploi at the same date. This information is available at the region (3) × aggregated FAP code (87) level.

- `tension_flow_PE`: the ratio between the *flow* of vacancies posted at Pôle emploi during the last trimester of 2015 and the *flow* of "catégorie A" job seekers newly registered in "catégorie A" at Pôle emploi during the same period. This information was the original indicator disclosed by the DARES.

- `tension_index_PE`: an index summarising nine indicators related to tightness on the labor market. This index has already been used by Signorelli (2019) in a paper studying the impact of a French reform facilitating the access to some labor markets in high tightness for migrants. The French commission that selected those markets stated that it used those nine indicators, hence the choice of Signorelli (2019) to build a synthetic index accounting for all of them. The indicators considered are the following: i) the ratio between flows of vacancies/job supply and job seekers/job demand registered during the reference period, ii) the stock of job supply, iii) the stock of job demand, iv) the evolution in the stock of demand and supply, v) the turnover rate of job seekers at the end of the month, and vi) the share of long term contracts within the job offers. A z-score of all these indicators in built, and the index is the unweighted sum of these z-score. As for the previous indicator, this information is available at the region (3) × FAP code (87) level.

Local labor market tightness data from 2016 BMO survey The BMO survey has been conducted since 2015 by PE in order to improve their knowledge about local labor market tightness. It surveys firms in different commuting zones (*bassins d'emploi*) about the number of hirings they schedule for the following year for different occupations — using FAP as occupational classification[25]. Additionally, it asks whether some of these hirings, in some occupations, are expected to be "difficult" because of either a lack of applicants or a lack of skills among the applicants pool. This BMO information gives us an idea about the level of labor market tightness as it informs us (i) about the raw numerator of tightness — through the measure of

[25] For a description of the BMO survey and the results published by PE based on it, see: `http://www.pole-emploi.org/accueil/actualites/infographies/difficultes-de-recrutement-pourquoi-certains-metiers-sont-en-tension.html?type=article`. For access to BMO survey data for all available years (from 2015 onwards), see: `https://statistiques.pole-emploi.org/bmo/`

total number of hirings per market — and (ii) about the tightness indicator itself, as difficult hirings are synonymous with a low number of (skilled enough) applicants with respect to the number of vacancies that need to be filled, i.e. a high labor market tightness. For the definition of markets here, I decided to use as geographical units departments (17 levels) and aggregated FAP codes (87 levels) as occupational units, as it seemed to me that this was defining not to incredible markets while still leaving enough observations per markets[26]. Again, as evoked in the previous paragraph, I must admit that there is room for improvement in the definition of markets here.

List of the different indicators based on BMO survey with description:

- `tension_nb.recrut.diff_BMO`: number of difficult hirings scheduled in the following year (2016) as reported by firms in the BMO survey. Hirings can be deemed "difficult" (according to BMO documentation) either because of a lack of skills in the applicants pool, or just because of a lack of applicants. This information is available at the département (17) $\times$ FAP code (87) level[27].

- `tension_nb.recrut.diff_BMO_REGLEVEL`: same than the previous variable, but at the region (3) $\times$ FAP code (87) level, for robustness check (since regions are the geographical level used for indicators extracted from STMT data).

- `tension_ratio.recrut.tot_BMO`: ratio between the number of hirings scheduled in the following year (2016) reported in BMO survey and the number of job seekers in the corresponding market. This information is available at the département (17) $\times$ FAP code (87) level[28].

- `tension_ratio.recrut.tot_BMO_REGLEVEL`: same than the previous variable, but at the region (3) $\times$ FAP code (87) level (robustness check).

[26]Moreover, I did not have in my data set extracted from PE administrative records the information about the *bassin d'emploi* in which each job seeker was located — the finer geographical information I had in this data set was actually departments.

[27]This data is even available at a finer geographical level ("Bassins d'emploi"), but I did not know, in the dataset extracted from the FH of Pôle emploi, the "bassin d'emploi" of each job seeker. The finer geographical level for which I had information in the dataset was the département, hence my choice.

[28]I did not know the number of job seekers by département $\times$ FAP code (and not even the number of job seekers by département). I had to estimate the number of job seekers by département (based on measures of the number of individuals by département, localised unemployment rates, and localised participation rates. Then I assumed that the share of job seekers in each FAP code in a département were the sames than the shares at the region level, which I can determine using the STMT data to which I have access.

- `tension_ratio.recrut.diff_BMO`: same as `tension_ratio.recrut.tot_BMO`, but using the number of *difficult* hirings scheduled in the year.

- `tension_ratio.recrut.diff_BMO_REGLEVEL`: same than the previous variable, but at the region (3) $\times$ FAP code (87) level (robustness check).

- `tension_sh.recrut.diff_BMO`: share of difficult hirings among all hirings scheduled within the year — information is available at the département (17) $\times$ FAP code (87) level.

3.2.3 Descriptive statistics and balancing tests

As mentioned earlier, data from PE administrative records include an important set of sociodemographic: gender, age, highest level of education completed, family status (married, widowed etc.), number of children, former kind of occupation (blue collar worker, manager etc.), nationality, region of living[29]. Along with these variables, I have also access to information about the search process of individuals: occupational code (ROME classification) of the job desired, reason for unemployment (personal or economic layoff, resignation etc.), years of experience in the job desired, type of contract desired (full/part time, permanent/fixed-term etc.). Table 1 reports summary statistics of a large number of these variables, as well as balancing tests results. First two columns present the average value of each covariate respectively among control and treated units, while the third and last column reports the balancing test result for each variable.

Looking at the first two columns of Table 1 reveals substantial heterogeneity in the characteristics of the job seekers involved in this experiment: there are around 24% of individuals that have some college education or more (BAC + 3 years or more), but there are also 38% of individuals that have an education level below BAC level (high school dropouts and vocational). A large share of individuals (45%) search in a job in which they have more than 5 years of experience, but a significant proportion (16%) desire a job in which they have no experience at all. Most individuals were previously working as clerical workers (64%), but there are also 15% of the job seekers involved in the experiment that used to work as managers.

The last column of Table 1 presents balancing tests. Reassuringly, for most of the covariates tested, the null hypothesis of equality between control and treated means fails to be rejected, which suggests that the randomisation of treatment assignment was properly executed.

[29]It also includes the number of the *agence locale pour l'emploi* (ALE) in which the job seeker is registered, which includes the department number. This allows me to have this extra information about th localization of each job seeker.

Table 1: Balancing tests

Variable name	Control mean (1)	Treated mean (2)	(1) = (2)
Sociodemographic variables			
Female	0.490	0.489	0.603
French	0.846	0.847	0.805
European	0.045	0.047	0.178
African	0.089	0.087	0.173
Other nationality	0.018	0.018	0.892
Aged below 25	0.145	0.145	0.844
Aged 26 to 35	0.273	0.273	0.951
Aged 36 to 45	0.227	0.229	0.540
Aged 46 to 55	0.198	0.199	0.696
Aged above 56	0.156	0.155	0.383
No exp. in the job	0.161	0.160	0.887
1 to 5 years of exp in the job	0.384	0.386	0.326
More than 5 years of exp in the job	0.455	0.453	0.392
No child	0.570	0.567	0.265
One child	0.179	0.180	0.514
More than one child	0.252	0.253	0.487
Education below BAC (including vocational)	0.383	0.386	0.297
BAC	0.228	0.227	0.373
Short higher education (BAC + 2 y.)	0.147	0.145	0.235
Some college education (BAC + 3 to 4 y.)	0.110	0.110	0.718
Graduate (BAC + more than 5 y.)	0.131	0.133	0.238
Blue collar	0.113	0.115	0.313
Clerical worker	0.648	0.645	0.198
Technician	0.091	0.092	0.554
Manager	0.147	0.148	0.726
Ile-de-France (Paris region)	0.694	0.693	0.585
Pays-de-la-Loire	0.172	0.173	0.633
Lorraine	0.134	0.134	0.834
Married	0.423	0.424	0.453
Search for a permanent contract	0.931	0.934	0.075
Search for a fixed-term contract	0.061	0.059	0.068
Search for seasonal work	0.007	0.007	0.892
Search for a full-time position	0.880	0.882	0.161
End of maternity/sick leave	0.046	0.048	0.069
End of contract	0.232	0.229	0.185
Economic layoff	0.041	0.041	0.951
Personal layoff	0.173	0.173	0.902
Resignation	0.018	0.017	0.265
Other reasons of unemployment	0.490	0.492	0.447
Labor market tightness information			
tension_stock_PE	0.050	0.049	0.078
tension_flow_PE	0.497	0.493	0.070
tension_index_PE	0.038	0.011	0.040
tension_nb.recrut.diff_BMO	317.223	320.291	0.252
tension_nb.recrut.diff_BMO_REGLEVEL	1,728.162	1,725.810	0.796
tension_ratio.recrut.tot_BMO	0.779	0.782	0.446
tension_ratio.recrut.tot_BMO_REGLEVEL	0.381	0.381	0.715
tension_ratio.recrut.diff_BMO	0.264	0.266	0.275
tension_ratio.recrut.diff_BMO_REGLEVEL	0.130	0.130	0.793
tension_sh.recrut.diff_BMO	0.335	0.337	0.113
Experiment information			
Opened LBB e-mail	0	0.353	0
Followed LBB link	0	0.190	0

Note: The first and second columns report the mean value of variables over the sample of job-seekers assigned respectively to control group and treatment group (those receiving the e-mail), who were still in "catégorie A" when the e-mails were sent on December 15$^{\text{th}}$ and for which there are no missing values (149,286 individuals). The third and last column report the p-value for the t-test of equality between control mean and treated mean.

4 Model

The aim of this section is to offer different ways of modelling the effect of the treatment studied — i.e. sending an e-mail incentivizing the use of LBB platform — on job search and the job finding probability of job seekers. I begin with a very simple job search model with exogenous search effort: this simple framework already allows to consider various potential cases. I then turn to a job search model with endogenous search effort in an attempt to provide some extra understanding of the mechanisms that could be at play.

Unfortunately, I do not have enough information in my data — in particular, not enough statistical power and not enough (not at all) information about job search intensity — to bring any structural model to the data. Several models presented here can be used equivalently to explain the same empirical observation, and I will not be able to choose between those in my discussion of results in the following sections. However, I still view this theoretical exercise as useful as it allows to keep in mind the different mechanisms that could be at play to explain the results that I will present in Section 5.

Section 4.1 builds upon a simple model of job search with exogenous searche intensity. Section 4.2 uses as a starting point a model with endogenous search effort, and has the main goal of proposing one candidate explanation for an heterogeneous threat effect on search intensity with respect to labour market tightness. It builds on the idea of biased beliefs (optimism) of job seekers (Spinnewijn, 2015) that would be (partially) erased by the mere fact of receiving a PE e-mail — as if this e-mail acted as a call for consistency of beliefs. This last section is developed has the mere purpose of offering one possible conceptual framework for such an heterogeneous treatment effects, and without claiming that this should necessarily be considered as the most relevant/plausible one.

4.1 A basic search model with exogenous search effort

4.1.1 Assumptions

Time is continous. There are different local labor markets, indexed by j. The number of workers on a each market is normalized to 1. The rate at which job seekers and firms meet is described by a constant returns to scale matching function $m(.)$, assumed common across markets. Let u_j denote the measure of unemployed workers (job seekers[30]) in market j. Let

[30] This model will consider only the case of job search from unemployed workers (no on-the-job search).

v_j denote the measure of vacancies and $\overline{s}_j$ job seekers' average search intensity on this market. The flow of matches per unit of time in this market is given by $m(\overline{s}_j u_j, v_j)$, increasing in both arguments and with $m_{11} < 0$ and $m_{22} < 0$[31]. Assuming that the matching function has constant returns to scale and is increasing and concave in both arguments is relatively common in the literature. Empirical work has shown that labor market data is consistent with these hypotheses (see Pissarides (2000) and Petrongolo and Pissarides (2001) for surveys of these studies). Let $\theta_j \equiv \frac{v_j}{\overline{s}_j u_j}$ measure labor market tightness. A job seeker who searches with intensity s meets a firm at rate:

$$s\mu(\theta_j) = s\frac{m(\overline{s}_j u_j, v_j)}{\overline{s}_j u_j} = sm(1, \theta_j) \tag{4.1}$$

increasing in θ_j by assumption on $m(.)$, the matching technology.

N.B.: during this whole section, we are reasoning in partial equilibrium. This is justified by the fact that the treatment studied concerns a rather small proportion of job seekers on each market, and is dispersed over the country.

4.1.2 The effect of an exogenous shock on search effort ("threat effect")

A channel through which our treatment could have an effect on the job finding rate of job seeker could be the "threat effect" of the e-mail sent (independently from the content of this e-mail, that we consider as having no effect for now). Indeed, it could be that receiving an e-mail from the public placement agency is interpreted by job seekers as a reminder of the fact that they are expected to search for jobs as a counterpart for their unemployment benefits. If job seekers have the feeling that they do not search enough, and that this e-mail sounds as a call to order from the public placement agency, then the e-mail can be in itself an exogenous shock on their search effort. Let $p(s, \theta) \equiv s\mu(\theta_j)$ denote the rate (or probability per unit of time) at which a job seeker who searches with intensity s meets a firm in this model. From equation 4.1, and assuming a "threat effect" on s that would not depend on θ, we have:

$$\frac{\partial p(s, \theta_j)}{\partial s} = \mu(\theta_j)$$

increasing in θ_j by assumption on the matching technology. This means that in our model, such a "threat effect" would be heterogeneous with respect to the labor market tightness: the higher θ_j, the higher the effect of the e-mail on the job finding rate.

[31] I denote by f_1 the first derivative of f w.r.t. its first argument, and f_{11} the second derivative of f w.r.t. its first argument. Similarly, f_x denote the first derivative of f w.r.t. x, and f_{xx} the second derivative of f w.r.t. x.

However, the "threat" effect of the PE e-mail on s is not necessarily homogeneous with respect to θ_j. A more general case would be to model the threat effect as $T \in \{0, 1\}$ such that we would have the effect of T on $p(s, \theta_j, T)$ equal to:

$$p(s, \theta_j, T = 1) - p(s, \theta_j, T = 0) = \underbrace{[s(T = 1) - s(T = 0)]}_{\Delta s} \cdot \mu(\theta_j)$$

and the derivative of this effect with respect to θ would be:

$$\frac{\partial p(s, \theta_j, T = 1) - p(s, \theta_j, T = 0)}{\partial \theta_j} = \frac{\partial \Delta s}{\partial \theta_j} \cdot \mu(\theta_j) + \Delta s \cdot \mu_{\theta_j}(\theta_j)$$

In this case, the heterogeneity of the "threat effect" with respect to θ_j can goes in all different ways: it only depends on the elasticities of Δs (the threat effect on search intensity s) and $\mu(\theta_j)$ (the matching rate between applications and vacancies) with respect to $theta_j$ (see proof 1). Indeed, in the case in which $\frac{\partial \Delta s}{\partial \theta_j} < 0$, meaning that the threat effect on search intensity would be lower when tightness is higher, it can be that the overall threat effect on the job finding probability would be decreasing: the condition is simply that the absolute value of the elasticity of the threat effect on s w.r.t. θ_j (which would be negative) is larger than the elasticity of the matching rate of applications with vacancies w.r.t. θ_j (which is positive by assumption).

Result 1 In a job search model with exogenous search effort, the mear fact of receiving an e-mail from the public placement agency could have a (positive) "threat effect" on the job finding probability, through an effect on search intensity. If this effect on search intensity does not depend on labor market tightness, then the overall effect on job finding probability should be increasing in labor market tightness (because of the assumptions made on the matching technology). Yet if for some reason this effect does depend on tightness, and if in particular it is lower and lower as tightness increases, then it can be that the overall effect on job finding probability is decreasing in tightness. Notice that Section 4.2 has the main purpose of proposing one candidate explanation for why it could be the case under some conditions.

4.1.3 The effect on the matching function

We now turn to modelling the effect of the content of the e-mail, i.e. the job search assistance provided by the online platform LBB. There are two main ways of understanding hwo this help works. The interesting fact is that these two different interpretation lead to a different relationship between the effect and labor market tightness.

The online platform as a vacancy provider The first way to look at the online job search assistance provided by LBB is to consider it as a tool that allows job seekers using it to have access to *more vacancies*. This is in a way the vision behind the original hope of economists about the development of the Internet and its impact on the labor market — see Section 1 for a brief review of this literature. Indeed, from the workers' perspective, the original view was to see the Internet as a way to access to much more vacancy postings through job boards that would gather them and offer an easy way to compare them.

If this is what LBB does, then the treatment goes through an increase of the vacancies in the matching function for individuals treated, i.e. an increase of θ_j faced by individuals: $\Delta\theta_j = \theta_j(\tau = 1) - \theta_j(\tau = 0)$[32]. The effect on the job finding probability is given by — considering that s is indeed exogenous and therefore not affected by the change in v_j:

$$\begin{aligned} p(s, \theta_j, \tau = 1) - p(s, \theta_j, \tau = 0) &= s \cdot [\mu(\theta_j, \tau = 1) - \mu(\theta_j, \tau = 0)] \\ &= s \cdot [\mu(\theta_j + \Delta\theta_j) - \mu(\theta_j)] > 0 \end{aligned}$$

Given the usual assumptions on the matching technology that imply the concavity of $\mu(.)$, we have that the derivative of this effect with respect to the labour market tightness θ_j is *negative*. Hence in a job search model with exogenous search effort, and if online job search assistance plays mainly the role of a vacancy provider, then the treatment effect should be — ignoring "threat effects" — (positive and) *decreasing in labour market tightness*.

The online platform as a vacancy sorter Yet one could understand the help provided by LBB in another way: a job seeker using the platform is finding more often the relevant firms to which they should send their applications. This is actually the way LBB staff think about their action. One might consider this as the new role of online platforms in job search: nowadays, everybody can look for vacancies on the Internet on a plurality of job boards; but there is room for online job search assistance if some platforms offer a way to select the more appropriate vacancies. This would actually be consistent with the observations of Bren***cic*** (2014), underlining that the number of postings that visitors of a job board review on average is not correlated with the number of available postings on the job board, and represents a small

[32] N.B.: This is obviously dependent on the fact that we are reasoning in partial equilibrium here.

fraction of all postings on the site.

If using LBB is about sorting vacancies and finding the right ones on which to focus, then the treatment can basically be viewed as an increase in the probability that when an application meets a vacancy, it ends up as an hiring. In other words, it can be modelled as an increase in aggregate matching efficiency. Let us now assume that the flow of matches per unit of time in this market is no longer $m(\overline{s}_j u_j, v_j)$, but $m_0(\tau) \cdot m(\overline{s}_j u_j, v_j)$ where $m_0(\tau)$ is the aggregate matching efficiency. $m_0(\tau)$ can also be understood as the probability that a "match" between a vacancy and an application — which probability is given by $m(\overline{s}_j u_j, v_j)$ — becomes a new hiring. It is shifted upward when τ — the indicator for being encouraged to use LBB — is equal to 1. We therefore have:

$$p(s, \theta_j, \tau) = s \cdot \frac{m_0(\tau) \cdot m(\overline{s}_j u_j, v_j)}{\overline{s}_j u_j} = s \cdot m_0(\tau) \cdot m(1, \theta_j) = s \cdot m_0(\tau) \cdot \mu(\theta_j)$$

"Threat effects" aside, the effect of the LBB e-mail on the job finding rate goes through an increase of m_0 for treated individuals, and is given by:

$$p(s, \theta_j, \tau = 1) - p(s, \theta_j, \tau = 0) = s \cdot \mu(\theta_j) \cdot (m_0(1) - m_0(0))$$

increasing in θ_j by assumption on the matching technology. This means that in our model, if LBB plays mainly a role of vacancy sorter, then the treatment effect — threat effects aside — would be (positive and) *increasing in labour market tightness.*

Result 2 In a job search model with exogenous search effort, there are two main ways to see the effect of an online platform such as LBB.

i. If online job search assistance plays mainly the role of a *vacancy provider*, then the treatment effect should be — "threat effects" aside — (positive and) *decreasing in labour market tightness.*

ii. If online job search assistance plays mainly the role of a *vacancy sorter*, then the treatment effect should be — "threat effects" aside — (positive and) *increasing in labour market tightness.*

4.2 A search model with endogenous search effort

The model presented here is borrowed from Shimer (2004), that offers one way of modelling endogenous search effort. I extend it in three ways. First, I allow for multiple independent markets. Second, I introduce biased beliefs of job seekers regarding the probability of acceptance of their applications — in the spirit of the work of Spinnewijn (2015) and Mueller et al. (2019), that document this kind of biases among job seekers in the US. Second, I model the effect of offering job search assistance through an online platform such as LBB, considering that this increases the probability of acceptance of applications from a job seeker using the platform. I comment these extensions in more details below, after having presented the basic model and derived the original formula for endogenous search effort. Notice that I will not develop on the effect of these extensions on the equilibrium value of search effort; this would not be of great interest for interpreting the results I observe in my data, since the experiment at hand was small and spread in the territory, hence it is reasonable to believe that I cannot detect equilibrium effects.

4.2.1 Assumptions

Time is continous. There are different local labor markets, indexed by j. There are two types of agents, workers and firms, that are risk-neutral and discount future payoffs at rate $r > 0$. The number of workers on a each market is normalized to 1. A job seeker, in order to generate s efficient search units[33], has to devote hs^ν hours to job search with $h > 0$ and $\nu > 1$. This function is increasing and convex in s. A way to justify this assumption is that if job seekers start their search with the most accessible vacancies, one can reasonably assume that the time needed for an extra application is increasing with the number of applications already sent, hence the convexity of the function. Normalizing the total number of hours available to a job seeker to 1, it means that the leisure time available to a job seeker searching with efficiency s is given by $1 - hs^\nu$. Let z represent the imputed value of leisure if a job seeker allocate all his/her time to leisure. Then we can specify imputed income during unemployment (not considering unemployment benefits in the model) as $z(1 - hs^\nu)$, and the cost of job search as zhs^ν.[34] An employed worker receives a wage w_j determined by Nash bargaining between the firm and the

[33] In a continuous time model, this search effort is not necessarily easy to interpret. Yet one can still think about it in terms of the number of applications sent (though it does not need to be an integer).

[34] For the sake of simplicity, we are assuming that the utility of leisure is not diminishing on the margin. Otherwise, the imputed income from leisure would be a concave (instead of linear) function of the time allocated to leisure. This assumption is not crucial here for the mechanisms studied, and it simplifies the exposition.

worker when he/she is hired. Jobs are identical on a given market j, but end exogenously at rate $\lambda > 0$ (considered the same across all markets for simplicity), leaving the worker unemployed.

Firms can open (and/or maintain opened) vacancies by paying a flow cost c. When a job is filed, it generates a flow profit equal to $x_j - w_j$ (where x_j is the productivity of a worker on market j) until the job is "destructed", leaving the firm with nothing. Free entry on each market drives the value of posting a vacancy to zero.

Wages are determined by Nash bargaining between employed workers and firms. Let $E(w_j)$ denote the expected present value of lifetime income for a worker employed at a current wage w_j, U_j denote the expected present value of income for an unemployed worker, and $J(w_j)$ denote the expected present value of profit from a particular filled job. The wage w_j^* is set to maximize the weighted product of surplus from the match in excess of threat points (i.e. outside options) of each party, giving:

$$w_j^* = \arg\max_{w_j} (E(w_j) - U_j)^{\gamma} J(w_j)^{1-\gamma} \tag{4.2}$$

where γ denotes worker's bargaining power (and will be assumed constant across markets, again in order to simplify to exposition).

The rate at which job seekers and firms meet is described by a constant returns to scale matching function $m(.)$, assumed common across markets. Let u_j denote the measure of unemployed workers (job seekers[35]) in market j. Let v_j denote the measure of vacancies and $\overline{s}_j$ job seekers' average search intensity on this market. The flow of matches per unit of time in this market is given by $m(\overline{s}_j u_j, v_j)$, increasing in both arguments and with $m_{11} < 0$ and $m_{22} < 0$. Let $\theta_j \equiv \frac{v_j}{\overline{s}_j u_j}$ measure labor market tightness. A job seeker who searches with intensity s meets a firm at rate:

$$s\mu(\theta_j) = s\frac{m(\overline{s}_j u_j, v_j)}{\overline{s}_j u_j} = sm(1, \theta_j) \tag{4.3}$$

increasing in θ_j, and a firm meets a worker at rate:

$$\eta(\theta_j) = \frac{m(\overline{s}_j u_j, v_j)}{v_j} = m\left(\theta_j^{-1}, 1\right) = \frac{\mu(\theta_j)}{\theta_j} \tag{4.4}$$

[35]This model will consider only the case of job search from unemployed workers (no on-the-job search).

4.2.2 Derivation of optimal search effort

We can describe the economic environment for each agent through a combination of Bellman equations. For workers on market j, we have:

$$rU_j = \max_s \left\{ z\left(1 - hs^{\nu}\right) + s\mu(\theta_j)\left(E\left(w_j^*\right) - U_j\right)\right\} \tag{4.5}$$

$$rE(w_j) = w_j + \lambda(U_j - E(w_j)) \tag{4.6}$$

Equation (4.5) gives the flow value of payoffs for an unemployed worker. He/She chooses his/her search intensity s in order to maximize his/her leisure $z\left(1 - hs^{\nu}\right)$ plus the returns from search, that corresponds to the matching probability $s\mu(\theta)$ times the capital gain upon getting a job. Equation (4.6) gives the flow value of payoffs for an employed worker. He/She receives a wage rate w_j, but the job ends at rate λ, leaving the worker unemployed. For firms, we have relatively similar equations:

$$rV_j = -c + \eta(\theta_j)\left(J\left(w_j^*\right) - V_j\right) \tag{4.7}$$

$$rJ(w_j) = x_j - w_j - \lambda J(w_j) \tag{4.8}$$

Equation (4.7) gives the value of a vacancy, equal to the probability of meeting a worker times the resulting capital gain, minus the posting cost c. Equation 4.8 is the counterpart of equation 4.6 for firms, except that a firm is left with nothing when a job ends. Note that free entry drives the value of a vacancy to $V = 0$.

Injecting equations (4.6) and (4.8) in the Nash bargaining equation (4.2), we get:

$$\frac{E\left(w_j^*\right) - U_j}{\gamma} = \frac{J\left(w_j^*\right)}{1-\gamma} = \frac{x_j - rU_j}{r+\lambda} \tag{4.9}$$

Substituting this into equations (4.5) and (4.7) yields:

$$rU_j = \max_s \left\{ z\left(1 - hs^{\nu}\right) + \frac{s\mu(\theta_j)\gamma(x_j - rU_j)}{r+\lambda}\right\} \tag{4.10}$$

$$c = \frac{\eta(\theta_j)(1-\gamma)(x_j - rU_j)}{r+\lambda} \tag{4.11}$$

once the second equation has been simplified using the free entry condition that implies $V = 0$. From there, it is possible to combine equation (4.10) and (4.11) using the equality $\eta(\theta_j) = \frac{\mu(\theta_j)}{\theta_j}$ derived in equation (4.4), implied by the constant returns assumption made on the matching function $m(.)$. We get:

$$rU_j = \max_s \left\{ z\left(1 - hs^{\nu}\right) + \frac{s\theta_j c\gamma}{1-\gamma} \right\} \tag{4.12}$$

Assuming an interior solution, we can derive from this equation the first order condition for the choice of search intensity s^* of a worker on market j. We have:

$$\nu h s_j^{*\nu-1} = \frac{\theta_j c\gamma}{z(1-\gamma)} \tag{4.13}$$

$$\Leftrightarrow \quad s_j^* = \left(\frac{\theta_j c\gamma}{\nu h z(1-\gamma)}\right)^{\frac{1}{\nu-1}} \tag{4.14}$$

Notice that since for now we assumed that all workers are identical, we have $s_j^* = \overline{s}_j$. Since $\nu > 1$, equation (4.14) implies that search effort is increasing with labor market tightness θ_j. The intuition behind this relationship is that applications from individuals on markets with a larger labor market tightness have a higher probability of being matched with a vacancy (since $\mu(\theta)$ is increasing in θ). Therefore, the expected returns from search are higher on markets with a higher θ, and job seekers respond to this higher incentive by searching more intensively.

4.2.3 Biased beliefs

In such a model with search effort being the result of an optimization, one can wonder how some "threat effects" as the ones described in section 4.1.2 could still be at play. What could be the mechanisms through which an e-mail from the public agency could in itself change the choice of search intensity of job seekers?

I build on the articles by Spinnewijn (2015), Mueller et al. (2019) and Altmann et al. (2018) in order to offer a way of modelling this. Based on surveys, Spinnewijn (2015) and Mueller et al. (2019) document the fact that job seekers in the US tend to overestimate their job finding probability. More precisely, they show that this bias tend to be higher for the long-term unemployed. One candidate explanation for this is that job seekers may have "motivated beliefs", i.e. they would adjust their expectations to preserve their self-esteem and/or to increase their expected payoffs through optimistic expectations (see Brunnermeier and Parker (2005) and/or Koszegi (2006) for models of optimistic expectations). The mechanism would be the following: as unemployment last for longer and longer, the demand of optimism is larger and larger.

In the framework of the model from Shimer (2004) presented above, I introduce this notion of biased beliefs as a perception bias on the matching rate of one's applications with vacancies on a market. Let us assume that the matching rate perceived by job seekers is given by:

$$\hat{\mu}(\theta_j) = \mu(\theta_j)\varepsilon(\theta_j)$$

$$\text{with } \forall\theta_j > 0, \qquad \varepsilon(\theta_j) > 1$$

$$\varepsilon_{\theta_j}(\theta_j) < 0$$

The assumptions made on the function $\varepsilon(.)$ can be understood as follows: (i) the perception bias of job seekers is always towards optimism (i.e. they always *over*-estimate the rate at which their applications are matched with vacancies), and (ii) this bias is decreasing with labor market tightness. The second assumption comes directly from the observations of Mueller et al. (2019) who document that biases regarding the job finding probability is larger for the long-term unemployed. As in our model, the average duration of unemployment on a given market j is given by $\frac{1}{s\mu(\theta_j)}$ (decreasing in θ_j), we have that the representative job seeker on a market has been unemployed for a longer duration when labor market tightness on this market is higher. As a consequence, if we follow the observations by Mueller et al. (2019) that describe a perception bias increasing with unemployment duration, it seems reasonable to assume that the representative job seeker on a market has larger biases when labor market tightness on this market is higher. We will add an extra assumption on this function: $\varepsilon(.)$ is such that $\hat{\mu}(\theta_j)$ is still increasing in θ_j, just as $\mu(\theta_j)$ is. This simply means that the perception bias of job seekers is not large enough for the representative job seeker to believe that as labor market tightness is higher, the (perceived) matching rate of their applications with vacancies would be smaller. To ensure that this is not the case, we need a condition on the first derivative of $\varepsilon(.)$, which can be proven to be (see proof 2 in appendix B):

$$\forall\theta_j > 0, \quad \varepsilon_{\theta_j}(\theta_j) > -\frac{\mu_{\theta_j}(\theta_j)\varepsilon(\theta_j)}{\mu(\theta_j)}$$

One can see in Figure 1 a sketch of the functions $\mu(\theta)$ and $\hat{\mu}(\theta)$.

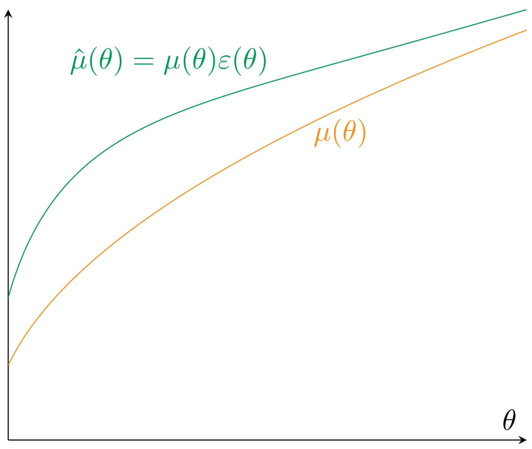


Figure 1: Biased beliefs as function of labor market tightness θ

The perceived matching probability of application to vacancies will be the one used by job seekers in their choice process. As a result, equation (4.10) in the model presented above becomes:

$$rU_j = \max_s \left\{ z\left(1 - hs^\nu\right) + \frac{s\hat{\mu}(\theta_j)\gamma(x_j - rU_j)}{r+\lambda} \right\} \tag{4.15}$$

We also have a new identity between the rate at which firms meet a worker, given by $\eta(\theta_j)$, and the *perceived* rate at which applications meet vacancies, which is given by:

$$\eta(\theta_j) = \frac{\mu(\theta_j)}{\theta_j} = \frac{\hat{\mu}(\theta_j)}{\theta_j \varepsilon(\theta_j)} \tag{4.16}$$

As a result, equation (4.12) of the model presented above becomes:

$$rU_j = \max_s \left\{ z\left(1 - hs^\nu\right) + \frac{s\theta_j\varepsilon(\theta_j)c\gamma}{1-\gamma} \right\} \tag{4.17}$$

From there, we can derive the new first order condition for the choice of search intensity s^* of a worker on market j, which is:

$$\nu h s_j^{*\nu-1} = \frac{\theta_j\varepsilon(\theta_j)c\gamma}{z(1-\gamma)} \tag{4.18}$$

$$\Leftrightarrow \quad s_j^* = \left(\frac{\theta_j\varepsilon(\theta_j)c\gamma}{\nu h z(1-\gamma)}\right)^{\frac{1}{\nu-1}} \tag{4.19}$$

The first order derivative of this function w.r.t. θ_j is:

$$\frac{\partial s_j^*}{\partial \theta_j} = \left[\varepsilon(\theta_j) + \theta_j\varepsilon_{\theta_j}(\theta_j)\right] \cdot \frac{c\gamma}{zh\nu(1-\gamma)} \cdot \frac{1}{\nu-1} \cdot \left(\frac{\theta_j\varepsilon(\theta_j)c\gamma}{zh\nu(1-\gamma)}\right)^{\frac{1}{\nu-1}-1} \tag{4.20}$$

which can be shown to be positive given the assumptions made on $\mu(.)$ and $\varepsilon(.)$ (see proof 3 in appendix B). Hence we still have that the search effort is increasing with labor market tightness, but this relationship is attenuated by the fact that as labor market tightness is reduced, the biased beliefs of job seekers are reduced, hence they do not increase their search effort as much as what they would have done in the absence of bias. On the contrary, for a given level of labor market tightness, their search effort is higher in the presence of the perception bias — since the partial derivative of s_j^* w.r.t. ε is positive (see proof 4 in appendix B).

These derivations above allowed us to model biased beliefs in a job search model, and to link them to the choice of search effort. I modelled these beliefs since I think that they are a interesting candidate channel in order to explain the "threat effect" of an e-mail from public placement agencies in the framework of job search model with endogenous search effort. Indeed, one could think about this e-mail as a reminder for unemployed individuals that they are not the only ones caring about their search; the public placement agency is potentially monitoring that they search to the best of their capacities. As a consequence, if the origin of the bias is some demand for optimistic expectations, the e-mail tends to "kill" the optimism by creating a demand for consistency between the behavior of the job seeker and what can observe the placement agency regarding the labor market conditions. Therefore, the effect of the e-mail can be understood, if this story is true, as a reduction of the perception bias of job seekers receiving the e-mail. The most abstract way of modelling this treatment would be to introduce it as a second parameter of the function $\varepsilon(.)$, so that we would have:

$$\hat{\mu}(\theta_j) = \mu(\theta_j)\varepsilon(\theta_j, T)$$

$$\text{with } \forall\theta_j > 0, \forall T \in \{0,1\}, \qquad \begin{aligned} \varepsilon(\theta_j, T) &> 1 \\ \varepsilon_{\theta_j}(\theta_j, T) &< 0 \\ \varepsilon(\theta_j, 1) - \varepsilon(\theta_j, 0) &< 0 \\ \varepsilon_{\theta_j}(\theta_j, 1) - \varepsilon_{\theta_j}(\theta_j, 0) &\geq 0 \end{aligned}$$

where T represents the fact of receiving the e-mail. The first and second assumptions on $\varepsilon(.)$ are those that were already made and explained earlier. The third one indicates that the fact of receiving the e-mail is assumed to reduce the perception bias. Lastly, the fourth assumption indicates that this effect is decreasing in θ_j. This seems reasonable to assume since it simply means, in other words, that a job seeker receiving the e-mail re-adapt to a larger extent his/her beliefs when those beliefs are more biased in the first place. Notice that we will come back to the discussion of the plausibility of this assumption at the end of this subsection, right before paragraph **Result 3**.

This leads us to the following formula for the choice of search effort:

$$s_j^* = \left(\frac{\theta_j \varepsilon(\theta_j, T) c\gamma}{\nu h z(1-\gamma)}\right)^{\frac{1}{\nu-1}}$$

The effect of the treatment on search effort is given by:

$$\begin{aligned}
\underbrace{s_j^*(T=1,\ .\ ) - s_j^*(T=0,\ .\ )}_{(\Delta s_j^*)} &= \left(\frac{\theta_j \varepsilon(\theta_j, 1) c\gamma}{\nu h z(1-\gamma)}\right)^{\frac{1}{\nu-1}} - \left(\frac{\theta_j \varepsilon(\theta_j, 0) c\gamma}{\nu h z(1-\gamma)}\right)^{\frac{1}{\nu-1}} \\
&= \left(\frac{\theta_j c\gamma}{\nu h z(1-\gamma)}\right)^{\frac{1}{\nu-1}} \cdot \left(\varepsilon(\theta_j, 1)^{\frac{1}{\nu-1}} - \varepsilon(\theta_j, 0)^{\frac{1}{\nu-1}}\right)
\end{aligned}$$

This effect is positive since we assumed $\varepsilon(\theta_j, 1) < \varepsilon(\theta_j, 0)$. The derivative of Δs_j^* w.r.t. θ_j is negative (see proof 5 in appendix B). In other words, under the assumptions we made on $\varepsilon(.)$, the effect of the e-mail on search effort is lower in labor markets with a larger market tightness. Notice that these assumptions are consitent with various functional forms for $\varepsilon(.)$. One could be the following:

$$\begin{aligned}
& \varepsilon(\theta_j, T) = 1 + \epsilon(\theta_j) - \alpha T \\
\text{with } \forall \theta_j > 0, \quad & \epsilon(\theta_j) > 0 \\
& \epsilon_{\theta_j}(\theta_j) < 0 \\
& \alpha > 0
\end{aligned}$$

In this case, it would mean that no matter the level of labor market tightness, the e-mail reduces by the same amount in *absolute* terms the perception bias.

Another could be the following:

$$\begin{aligned}
& \varepsilon(\theta_j, T) = (1 + \epsilon(\theta_j)) \cdot (1 - \alpha T) \\
\text{with } \forall \theta_j > 0, \quad & \epsilon(\theta_j) > 0 \\
& \epsilon_{\theta_j}(\theta_j) < 0 \\
& \alpha > 0
\end{aligned}$$

In this case, it would mean that the e-mail reduces by the same amount in *relative* terms the perception bias. Both these functional forms are consistent with the assumptions we made for the more general function $\varepsilon(.)$.

Since we do not consider equilibrium effects here, the job finding rate (per unit of time) is given by $p(T, \theta_j) = s_j^*(T, \ . \ )\mu(\theta_j)$. The effect of the treatment on this probability is given by the difference $p(1, \theta_j) - p(0, \theta_j)$, and the derivative of this effect w.r.t. θ_j is:

$$\frac{\partial p(1, \theta_j) - p(0, \theta_j)}{\partial \theta_j} = \underbrace{\frac{\partial s_j^*(T=1, \ . \ ) - s_j^*(T=0, \ . \ )}{\partial \theta_j} \cdot \mu(\theta_j)}_{<0} + \underbrace{\Big(s_j^*(T=1, \ . \ ) - s_j^*(T=0, \ . \ )\Big) \cdot \mu_{\theta_j}(\theta_j)}_{>0}$$

We see here that unlike in the job search model with exogenous search effort, here we have that the derivative of "threat effect" of a placement agency's e-mail w.r.t. θ_j can be negative (to be compared with result 1). The condition for this to be the case is:

$$\frac{\partial p(1, \theta_j) - p(0, \theta_j)}{\partial \theta_j} < 0$$

$$\Leftrightarrow \quad \frac{\partial s_j^*(T=1, \ . \ ) - s_j^*(T=0, \ . \ )}{\partial \theta_j} < -\frac{\Big(s_j^*(T=1, \ . \ ) - s_j^*(T=0, \ . \ )\Big) \cdot \mu_{\theta_j}(\theta_j)}{\mu(\theta_j)}$$

This is trivial: if the effect of the e-mail on search effort is decreasing at a sufficient rate in labor market tightness θ_j, then it compensates the fact that the matching rate between applications and vacancies is increasing in θ_j, and therefore the overall effect of the e-mail on the job finding probability is decreasing in θ_j. How likely is it that this inequality would hold? I cannot test properly this using my data, since there were no job seekers who received an e-mail from the placement agency without the LBB link, thus I cannot distinguish both effects[36]. However, the work by Altmann et al. (2018) suggests that it can be the case. Indeed, in their experiment, they merely send (through the public placement agency) a brochure to German job seekers without any personal information, almost uniquely as a reminder that (ibid) "active job search is a key to success", and that "many people greatly underestimate the impact of their personal initiative". Yet, despite the apparent absence of crucial information about job search in this brochure, the authors find significant (positive) effects of the brochure on job finding probability, and

[36] I do have an information about whether or not job seekers followed the link towards LBB contained in the e-mail. But the very fact of following the link can be correlated with a plurality of (unobservable) factors that could affect as well the job finding probability, such as motivation, thus I still cannot use this information to test in a convincing way for this feature of the "threat" effect.

more particularly when looking at its effects on the *long-term unemployed.* If we are willing to consider that their brochure does no contain much more than an empty e-mail from a placement agency, and that the channel through which the effect they observed is the one described in the framework I presented in this section, then it would mean that in their experiment, the inequality presented above did hold. Indeed, the fact that they find that the effect is larger among long-term unemployed would imply that on average, it would be larger on markets with a higher labor market tightness — since a higher θ_j would imply a larger share of long-term unemployed individuals.

Result 3 In a job search model with endogenous search effort and biased beliefs on the matching rate between applications and vacancies, and under some general assumptions on the function generating the perception bias, we show that the effect of receiving an e-mail from the public placement agency (that reduces the perception bias) on job search effort is decreasing in labor market tightness. If this effect is sufficiently decreasing in θ_j (as compared to the increase of the matching rate in θ_j), the "threat effect" of the e-mail on the job finding probability can be decreasing in θ_j.

4.2.4 Modelling online platforms' effect: an increase in matching efficiency

We now turn to modelling the effect of the job search assistance provided by the online platform LBB, in this job search model with endogenous search effort. As in section 4.1.3, we will consider that using the platform allows a job seeker to increase the matching rate between his/her applications and vacancies on the market.

We come back to assuming that the flow of matches per unit of time in this market is no longer $m(\bar{s}_j u_j, v_j)$, but $m_0(\tau) \cdot m(\bar{s}_j u_j, v_j)$ where $m_0(\tau)$ is a shift parameter for the matching rate of applications to vacancies, that is shifted upward when τ (the indicator for using LBB) is equal to 1. Let us make the following normalization $m_0(0) = 1$ (to clarify the exposition). We will further make the extra assumption that the treatment is given to a small enough share of job seekers that do not interact with each other[37] so that the average m_0 on each market is equal to $m_0(0) = 1$. We therefore have the matching rate between applications and vacancies equal to:

$$\mu(\theta_j, \tau) = \frac{m_0(\tau) \cdot m(\bar{s}_j u_j, v_j)}{\bar{s}_j u_j}$$

[37]This goes back to the assumption that we do not consider general equilibrium effects of the treatment(s) here.

This implies that we have the following matching function identity:

$$\eta(\theta_j) = \frac{m_0(0) \cdot m(\overline{s}_j u_j, v_j)}{v_j} = \frac{\mu(\theta_j, \tau)}{m_0(\tau)} \tag{4.21}$$

As a result, equation (4.12) of the model presented in section 4.2.2 becomes:

$$rU_j = \max_s \left\{ z\,(1 - hs^\nu) + \frac{s\theta_j m_0(\tau) c\gamma}{1-\gamma} \right\} \tag{4.22}$$

From there, we can derive the new first order condition for the choice of search intensity s^* of a worker on market j, which is:

$$\nu h s_j^{*\nu-1} = \frac{\theta_j m_0(\tau) c\gamma}{z(1-\gamma)} \tag{4.23}$$

$$\Leftrightarrow \quad s_j^* = \left(\frac{\theta_j m_0(\tau) c\gamma}{\nu h z(1-\gamma)} \right)^{\frac{1}{\nu-1}} \tag{4.24}$$

The effect of the treatment on search effort is positive, and given by:

$$s_j^*(\tau = 1, \ . \ ) - s_j^*(\tau = 0, \ . \ ) = \left(\frac{\theta_j c\gamma}{\nu h z(1-\gamma)} \right)^{\frac{1}{\nu-1}} \cdot (m_0(1) - m_0(0))^{\frac{1}{\nu-1}} > 0$$

The derivative of this effect w.r.t. θ_j is:

$$\frac{\partial s_j^*(\tau = 1, \ . \ ) - s_j^*(\tau = 0, \ . \ )}{\partial \theta_j} = \frac{1}{\nu - 1} \cdot (\theta_j)^{\frac{1}{\nu-1}-1} \cdot \left(\frac{c\gamma}{\nu h z(1-\gamma)} \right)^{\frac{1}{\nu-1}} \cdot (m_0(1) - m_0(0))^{\frac{1}{\nu-1}}$$

This derivative is positive, which means that the effect of the platform on job search is increasing with labor market tightness. Now, if we turn to the effect on the job finding rate, given by:

$$p(\tau, \theta_j) = s_j^*(\tau, \ . \ )\mu(\theta_j, \tau) = s_j^*(\tau, \ . \ ) \cdot m_0(\tau) \cdot m(1, \theta_j)$$

The effect of the treatment on this probability is given by the following difference:

$$p(1, \theta_j) - p(0, \theta_j) = \Big(s_j^*(\tau = 1, \ . \ ) \cdot m_0(1) - s_j^*(\tau = 0, \ . \ ) \cdot m_0(0) \Big) \cdot m(1, \theta_j)$$

This is positive since we have both $s_j^*(\tau = 1, \ .\ ) > s_j^*(\tau = 0, \ .\ ) \geq 0$ and $m_0(1) > m_0(0) > 0$. The derivative of this effect w.r.t. θ_j is:

$$\frac{\partial p(1,\theta_j) - p(0,\theta_j)}{\partial \theta_j} = \frac{\partial s_j^*(\tau = 1, \ .\ ) \cdot m_0(1) - s_j^*(\tau = 0, \ .\ ) \cdot m_0(0)}{\partial \theta_j} \cdot m(1,\theta_j)$$
$$+ \left(s_j^*(\tau = 1, \ .\ ) \cdot m_0(1) - s_j^*(\tau = 0, \ .\ ) \cdot m_0(0) \right) \cdot m_{\theta_j}(1,\theta_j)$$

This derivative can be shown to be positive (see proof 6 in appendix B). This means that in this framework, the effect of LBB is unambiguously increasing in labor market tightness.

Result 4 In a job search model with endogenous search effort, if using an online platform such as LBB increases the matching rate between applications of treated job seekers and vacancies, then (i) the use of LBB increases job search effort, even more so in market with a larger market tightness, and (ii) it increases as well the job finding probability, again to a larger extent in markets with a larger market tightness.

5 Results: ATE and First Heterogeneity Analysis

This section presents the analysis of average treatment effect (ATE) and a parametric approach to heterogeneity analysis (using interaction terms or sub-sampling), exploring treatment effect heterogeneity along (i) standard covariates (gender, age, education) and (ii) labor market tightness.

5.1 ATE and heterogeneity along "standard" dimensions

In this section, I present the results of ATE analysis and heterogeneity across usual covariates (age, gender, education).

Table 2 reports the results for ATE estimation (with and without controls) at different time horizons. The reader should have in mind that by "treatment" here, I actually mean ITT[38]. The table reveals that there is no significant effect detected at any time horizon, without or with controls included in the regression. This is further confirmed by the graphical evidence presented in Figure 2, which presents the control and treated average job finding rate and the difference between those two rates (with a 90% confidence interval) for all different time horizons available — up until two years after the experiment. There is not any point in time at which there is a

[38] See Section 3.1 for justification of this focus on ITT analysis

significant treatment effect at the 10% level[39].

Given the very low "take-up" of the program — only 35% of treated job seekers opened the e-mail, and only 19% of them actually clicked on the LBB link — it might not be a surprise that we do not detect any significant effect: this might be the consequence of a lack of statistical power, given that we do not expect the true effect (if it exists) to be very large.

Table 2: Intention-to-treat (ITT) at different time horizons

	Dependent variable:					
	Job found after 6 m.		Job found after 12 m.		Job found after 24 m.	
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	0.002	0.002	0.001	0.0005	0.002	0.002
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Constant	0.226***	0.517***	0.342***	0.732***	0.460***	0.925***
	(0.007)	(0.017)	(0.008)	(0.018)	(0.008)	(0.017)
Controls	No	Yes	No	Yes	No	Yes
Observations	149,286	149,286	149,286	149,286	149,286	149,286

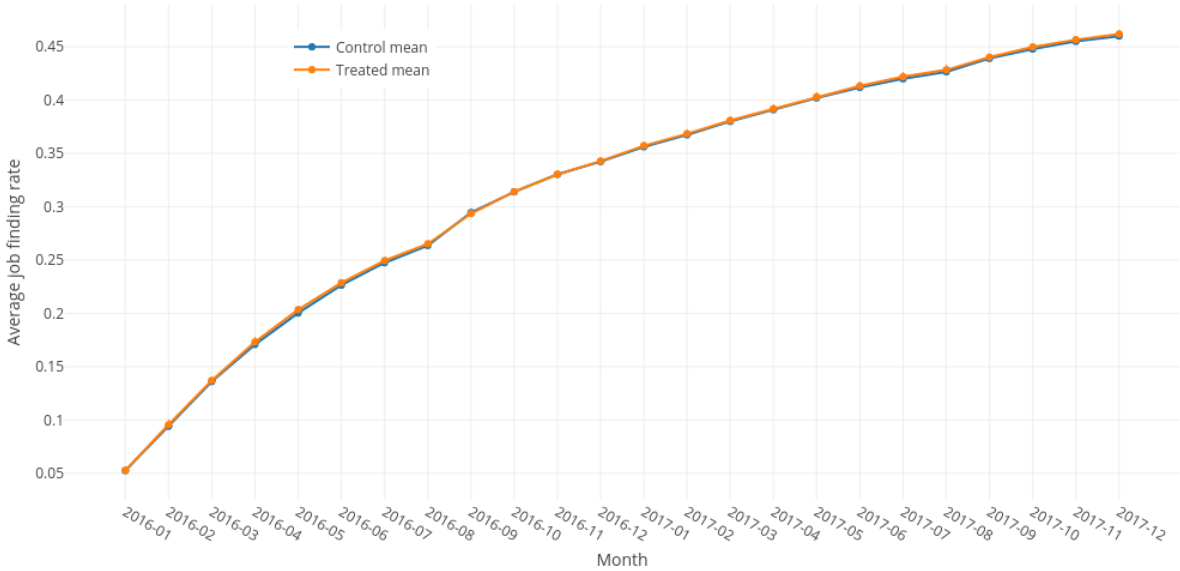
*p<0.1; **p<0.05; ***p<0.01

Note: Robust clustered standard errors are reported in parentheses (clusters are the FAP code × Employment agency (ALE) level). These regressions are performed over the whole sample of job seekers (without missing values). Controls include age, sex, level of education, experience in the job for which the job seeker if searching, and qualification (categorical variable indicating whether the individual is a blue collar, clerical worker etc.). The dependent variables are dummies indicating whether the individual has found a job x months after the beginning of the experiment (i.e. December 2015).

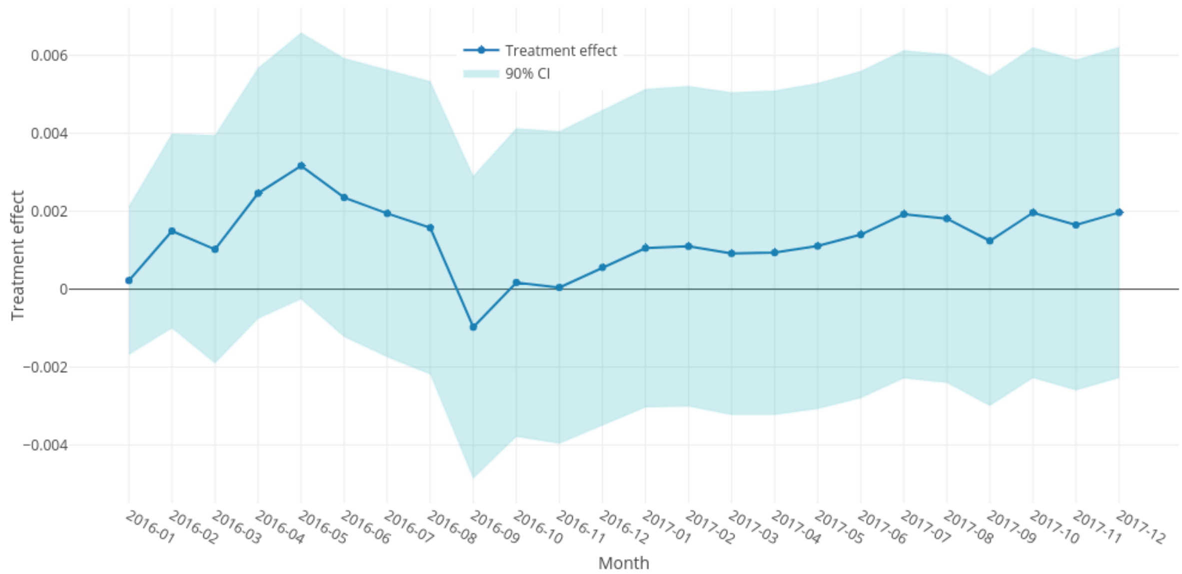
Given the absence of any detectable ATE, I decide to focus on heterogeneous treatment effects, in order to still learn (if possible) from this experiment despite this lack of statistical power. The hope is that there are certain subgroups in which the treatment has a large enough effect to be detected; if this is the case, then it is still a useful information to collect for the policy maker (and for the economist) as it might allow for a better targeting of the policy and/or a better understanding of the mechanisms at play.

In Table 3, I present some heterogeneity analysis with respect to usual covariates: gender, age and education. I run on each different subsample a simple regression with a dummy indicating whether the individual has found a job 6 months after the experiment as the dependent vari-

[39] I present in Table A.1 and in Figure A.1 similar results for the analysis of the effect of clicking on LBB link, instrumented by the treatment assignment variable — that should be taken with cautiousness given the concerns expressed in Section 3.1.



(a) Job finding rate, controls vs. treated



(b) ITT effect at different time horizons

Figure 2: Timing of treatment effect

Notes: In Figure 2a, we report average job finding rate for different time horizons, depending on the treatment status. In Figure 2b, we report average intention-to-treat (ITT) effect for different time horizons, which amounts to repeating the regression presented in Table 2 for all available dependent variables, from month = 1 to month = 24. Note that it corresponds more or less to the gap between the two curves presented in Figure 2a. The ITT effect for month "2017-01" is the effect of the ITT on the probability of having found a job between January 2016 and January 2017 (month = 12 in this case).

able, and a dummy for treatment status along with some controls as independent variables. I report the coefficient for the treatment status dummy, which corresponds the ATE estimates within each subgroup. For nearly all subgroups, I do not find any evidence of heterogeneous (not even significant) treatment effects. Yet, interestingly enough, I find that the treatment effect is significant (at the 5% level) for individuals with an level of education corresponding to college education or more (3 years or more after high-school diploma, BAC in french). This is interesting since it would suggest that individuals benefiting the most from this treatment are not those with the worse employment prospects — it might even be that it actually benefits to individuals with above median employments prospects. We will come back to this questioning in Sections 5.3 and 6.1.

Table 3: Heterogeneous treatment effects with respect to "classic" covariates

	Dependent variable:							
	Jod found after 6 months							
	Female	Male	Less 29	30-44	Above 45	Below BAC	BAC or Vocational	College or Graduate
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treatment	0.003	0.002	0.004	−0.001	0.004	0.001	−0.0002	0.008**
	(0.003)	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)
Constant	0.471***	0.550***	0.558***	0.469***	0.871***	0.458***	0.498***	0.636***
	(0.019)	(0.018)	(0.032)	(0.022)	(0.048)	(0.020)	(0.019)	(0.023)
Observations	79,999	76,287	39,493	53,703	56,090	57,768	56,009	35,509

*p<0.1; **p<0.05; ***p<0.01

Note: Robust clustered standard errors are reported in parentheses (clusters are the FAP code × Employment agency (ALE) level). These regressions are performed over subsamples of job seekers (without missing values) with characteristics described at the top of each column. Controls include age, sex, level of education, experience in the job for which the job seeker if searching, and qualification (categorical variable indicating whether the individual is a blue collar, clerical worker etc.). The dependent variable is a dummy indicating whether the individual has found a job 6 months after the beginning of the experiment (i.e. December 2015).

5.2 Interaction between labour market tightness and the treatment: a first exploration

In this section, I turn to the analysis of the interaction between the treatment effect and labour market tightness. As discussed while describing the data in Section 3, there is no standard indicator for labour market tightness. I therefore use different alternative measures of tightness, from different data sources.

I present in Table 4 the results of the parametric exploration of treatment effect heterogeneity with respect to the different tightness measures. For each measure, I build a dummy indicating whether the observation has a tightness above the sample median; I then run a regression with an interaction term between this dummy and the treatment assignment variable to look for heterogeneity[40].

Overall, what appears in this table is that we do not detect so much treatment effect heterogeneity with respect to any tightness indicator. The only indicator for which we do detect some significant heterogeneity is the one used in column (4), namely the one measuring the number of difficult hirings scheduled on the market, from BMO survey. One might notice that this is also the only regression in which I include the control for the total number of hirings scheduled on the market as measured by BMO. This is simply because the tightness indicator at hand in column (4) is not re-scaled to take into account the size of the market. That said, including this control in other regressions did not change the results substantially, hence this is not the reason for detecting a significant result in this specification.

The first main message I take out of this table is the decorrelation of the different tightness indicators collected: the sign of their correlation with the job finding probability (the outcome) is not even the same depending on the tightness indicator considered. This underscores the difficulty to come up with some unanimously recognized tightness indicator upon which we could build our studies. The second message is of course that if I were to interpret the only significant result of this table, it would suggest that the treatment effect is *increasing* in labour market tightness as measured by the number of difficult hirings scheduled on the market. However, given the lack of robustness of this result — not confirmed by any other tightness indicator — I am reluctant to over-interpret it for now.

5.3 Market job finding rate as a proxy for labor market tightness

This section adopts a clearly different approach than the one used in the previous one. Indeed, instead of trying to use tightness indicators coming from extra sources, I decided to build a last proxy for tightness based on my own original data source. The main intuition at the start of this strategy is the following: if we believe in the general equilibrium theory, then there is a simple link between the job finding probability and labour market tightness, given by:

$$p(\theta_j) = m(\bar{s}_j u, v_j) = m(\theta_j, 1)$$

[40] I present in Table A.2 the results when not discretizing the tightness indicators: there are no major changes qualitatively.

Table 4: Heterogeneous treatment effects with respect to various tightness measures

	Dependent variable:					
	Job found after 6 months					
	Stock ratio (PE)	Flow ratio (PE)	Index (PE)	Nb. diff. hirings (BMO)	Stock ratio (BMO)	Sh. diff. hirings (BMO)
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	0.003	0.001	0.0003	-0.001	0.003	0.004
	(0.003)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)
Above median tightness	0.014	-0.002	0.008	0.002	-0.004	0.026^{***}
	(0.010)	(0.008)	(0.010)	(0.008)	(0.007)	(0.009)
Tot. nb. hirings scheduled (BMO)				-0.00000^{***}		
				(0.00000)		
Treatment × Above median tightness	-0.001	0.003	0.004	0.007^{*}	-0.002	-0.004
	(0.004)	(0.005)	(0.004)	(0.004)	(0.004)	(0.005)
Constant	0.508^{***}	0.518^{***}	0.513^{***}	0.535^{***}	0.519^{***}	0.501^{***}
	(0.017)	(0.016)	(0.017)	(0.016)	(0.017)	(0.017)
Observations	149 286	149 286	149 286	149 286	149 286	149 286

*p<0.1; **p<0.05; ***p<0.01

Note: Robust clustered standard errors are reported in parentheses (clusters are the FAP code × Employment agency (ALE) level). These regressions are performed over the whole sample of job seekers (without missing values). Controls include age, sex, level of education, experience in the job for which the job seeker if searching, and qualification (categorical variable indicating whether the individual is a blue collar, clerical worker etc.). The dependent variable is a dummy indicating whether the individual has found a job 6 months after the beginning of the experiment (i.e. December 2015). The name of each column indicates the tightness measure used in the regression: the corresponding variable names are, in columns order, `"tension_stock_PE"`, `"tension_flow_PE"`, `"tension_index_PE"`, `"tension_nb.recrut.diff_BMO"`, `"tension_ratio.recrut.tot_BMO"`, `"tension_sh.recrut.diff_BMO"`. The significance test for the interaction term with the treatment variable (in bold) indicates whether or not there is treatment effect heterogeneity. In column (4), I included as control a proxy for the size of the market, namely total the number of hirings scheduled on the market (as measured by BMO), as the tightness indicator at hand in column (4) is not re-scaled to take this into account. Including this control in any of the 5 other regressions presented in this table was not changing the significance of the interaction term coefficient.

with $p(\theta_j)$ the average job finding rate on market j, $m(.)$ the matching function (increasing in θ_j), and θ_j the labour market tightness on market j.

Therefore, market job finding rate can constitute a proxy for labour market tightness. However, there is a technical issue to apply this using my data. Indeed, if I measure the average job finding rate (at 6 month) on each market using the whole sample of control units present in each market, then I am also computing the counterfactual outcome in the absence of treatment for each market. If there is any sampling error in the estimation of the average job finding rate per market, then once I will use this quantity and interact it with the treatment variable in a regression, the sampling error will be spuriously considered as heterogeneous treatment effect. If this is the case, then this will tend to bias my estimates in the following way: the higher the market job finding rate, the lower the treatment effect. In Table A.3, I report the results of such a regression. The fact that they are highly significant, of a large magnitude, and conclude precisely to a treatment effect in market job finding rate, is converging evidence of this kind of bias. In order to correct this bias, I build on Abadie et al. (2018) that proposes a way to correct for bias in randomized experiment studies when creating endogenous stratification based on a prediction of the outcome in the absence of treatment. This can be paralleled with what I am doing here. The way to correct such a bias is intuitively simple: one has to make some "honest" inference, meaning that the estimation of the outcome in the absence of treatment should be made using only on random part of the control units, while the other part would be used in treatment effect estimation. This way, one avoids the issue of sampling error being spuriously included in the interaction term coefficient estimated. Abadie et al. (2018) propose to repeat this procedure 100 times using 100 different splits of the control units. I add to this procedure an extra step, which is cross-fitting. Indeed, knowing that I already lack statistical power to detect treatment effect, I cannot afford the huge loss of statistical power potentially implied by honest inference. However, cross-fitting provides an alternative. I can estimate the market job finding rate using half of control units, and use this estimation for the other half of control units and half of treatment units; then I reverse the role of the sub-samples of control units, and use the new set of market job finding rate estimates for the last half of control units and treatment units missing an estimate. This way, I am no longer exposed to the bias caused by sampling error — as the estimates for market job finding rates for any control unit was computed without using this control unit — and I still keep all my observations for estimation of heterogeneous treatment effects.

Notice that instead of using the raw market job finding rate, I use a residualized measure: I take the residual of the market job finding rate as predicted by region (3 levels) and aggregated FAP codes (87 levels) fixed effects. That whay, when I discretize the measure between above vs. below median market job finding rate, the ranking is not affected by some regional or occupational particularities that could affect the matching function efficiency.

I report the results of this procedure in Table 5. The results are particularly interesting. Indeed, we observe that the sign of the interaction term coefficient is reversed, confirming our worries of sampling error biasing the results presented in Table A.3. Moreover, the treatment effect is found to be significant at the 10% level when using bootstrap standard errors based on 200 repetitions[41]. It would mean that the treatment effect is increasing in our proxy for labour market tightness: it is significantly positive on markets with a market job finding rate above median, and not significantly different from zero on markets with a market job finding rate below median. Notice that the magnitude of the coefficient is not that large: given that the treatment is found to be not statistically different from zero on markets with below median job finding rate, it would mean that the treatment increases by around half a percentage point the job finding probability of treated individuals on markets with an above median job finding rate. It is interesting enough to notice at this point that this result is both qualitatively and quantitatively in line with the result presented in column (4) of Table 4, using as tightness indicator the number of difficult hirings scheduled on each market as measured by BMO.

One might be worried that the heterogeneous effect documented here could be driven by different average level of education across markets that would change market job finding rate, even when residualized at the region $\times$ occupational level. It might still be the case since I residualize using aggregated FAP codes (87 levels)[42]. If it were the case, then markets with higher education would be more often classified with an above median average job finding rate; and as shown in Table 3, the treatment is found to be more efficient for individuals with college education or more. However, I should remind the reader that in the specification tested with the repeated slit sample estimator, I am controlling for the level of education of the individual. As such, I would argue that differential market average level of education is not likely to be at

[41] I acknowledge that a larger number of repetitions in order to determine standard errors would be appreciated. However, it had a large computational cost to repeat the repeated split sample estimator based on 100 splits for hundreds of time. One might want to take those results with caution because of this. In Abadie et al. (2018), they use bootstrap standard errors based on 1,000 repetitions — which is feasible since they are working on data sets far smaller than mine.

[42] This is still rather disaggregated. For instance, this distinguishes, in the restaurant industry, between bakers, cookers, maîtres d'hôtel and managers.

the origin of the heterogeneous effect documented here. Notice as well that I am controlling for region and aggregated FAP code fixed effects, meaning that this effect is identified across markets in different departments within the same region and the same occupational code. This is to take into account the possibility of a different matching function efficiency depending on geographical and/or occupational specificities.

That said, how can this result be related with the theoretical framework presented in Section 4.1 ? It is actually consistent with (at least) two alternative ways of modelling of the treatment. A first one is the basic "threat effect" of the PE e-mail, that would increase exogenously search effort independently of labour market tightness. This would result in a larger effect on the job finding probability for job seekers on markets with a high tightness; the intuition behind this result is that for these job seekers, a given increase in their search intensity (e.g. an increase of the number of applications sent) has major effects, because the matching rate between applications and vacancies is higher on markets in high tightness.

Yet these results are also consistent with another candidate explanation, in which LBB job search assistance acts as a *vacancy sorter*: it filters vacancies in order to help job seekers focusing on those that have a larger probability to actually match with them, and result in a hiring. This is comparable to an increase in aggregate matching efficiency for treated individuals, and it leads to a larger treatment effect on markets in high tightness — see Section 4.1 for further details. If this is what is really going on, then it is consistent with what LBB staff think about the role of their platform. It is interesting as the original hope regarding the Internet was mainly about a widening of job seekers' perspectives — and it is still something that authors like Belot et al. (2018b) have in mind — while these results would suggest another role, namely a better *targeting* of job seekers' effort. This might be of great importance knowing that this search effort is far from being unlimited — as documented by Bren**cic** (2014), who underlines that the number of postings that visitors of a job board review on average is not correlated with the number of available postings on the platform, and represents only a small fraction of all postings on the site.

Of course we cannot choose here between these two candidate explanation. A first step towards this choice would be to observe search intensity, so that we could study whether the treatment is actually driven by a modification/increase in job search intensity — which would be a bit more in line with the threat effect model[43]. On the contrary, if we were to find that search intensity is not modified by the treatment, then the second explanation might be deemed as the more relevant one.

Table 5: Heterogeneous treatment effects w.r.t. market job finding rate: "honest" repeated split sample estimator

	Dependent variable:
	Job found after 6 months
Treatment	(-0.0011) 0.0029
Above median market job finding rate (residualized)	0.0082** (0.0032)
Treatment × Above median market job finding rate (residualized)	0.0059* (0.0035)
Constant	0.4374*** (0.0238)
Controls	Yes

*p<0.1; **p<0.05; ***p<0.01

Note: Bootstrap standard errors based on 200 bootstrap repetitions. In order to avoid the issue of sampling error being spuriously included in the coefficient of the interaction term in the regressions presented in Table A.3, I measure the market job finding rate using half of the controls in each market — defined by departments (17 levels) × aggregated FAP codes (87 levels). I then take the residualized market job finding rate, which is the part of this rate that is not explained by some regional (3 different regions) and/or occupational (87 different FAP codes) features: this is my preferred indicator as one can argue that the matching function differs depending on geographical and/or occupational features. I then attribute the measured indicator to the other part of control units and to one (randomly selected) half of treated units. I then use cross-fitting in order to avoid losing too much observations: this means that I reverse the role of the subsamples of control units. Therefore, I end up with an estimate of the residualized market job finding rate for all observations. I then create a variable indicating whether this rate is above the median. I use this in the regression specified as in Table A.3 in column (4). I repeat this procedure 100 times with 100 different splits of the sample of control units, and I report the mean of the resulting estimates as the point estimate in the table. This procedure is inspired by Abadie et al. (2018) — I simply added cross-fitting.

[43]It would still not be a way to reject totally the possibility of an increase in matching efficiency. Indeed, as shown in Section 4.2, an increase in matching efficiency could be associated with a higher search intensity if search intensity if endogenous.

6 A Data-Driven Approach to Heterogeneity Analysis: What Can the Machine Tell ?

6.1 A first approach: focusing on individuals "at risk of long-term unemployment"

The approach used here is not the most data-driven, ML-focused one that I will implement in this section. However, I still decided to classify it in Section 6 because I attempt to use ML predictive power in order to implement endogenous stratification.

6.1.1 Heterogeneous treatment effects with respect to risk of long-term unemployment: using ML for endogenous stratification analysis

The first objective of this section is to replicate an heterogeneity analysis which is rather common in the literature of program evaluation, which is the study of treatment effect on individuals "most in need of help" — see Abadie et al. (2018) for a precise description of the usual strategies applied in order to do so, and Altmann et al. (2018) for an application in the context of an active labour market policy evaluation. The usual way to implement this kind of analysis is to use the whole sample of control units in order to estimate a basic linear model using baseline covariates to predict the outcome (here, whether a job is found 6 months after the experiment) in the absence of treatment. The predictions of this model for control *and* treated units are then used in order to study treatment effects for different intervals of predicted outcome in the absence of treatment. I implement this basic strategy using a basic linear probability model (LPM) in order to predict the outcome in the absence of treatment (see Table A.4 for the coefficients of the LPM), and report the heterogeneity analysis in Table A.5. There is not any significant treatment effect heterogeneity detected using this strategy.

Yet I tried to change a bit the approach to this problem, by modifying the model used in order to predict the outcome in the absence of treatment. I thought that I could use some ML method in order to build a better predictive model. I decided to use a regression forest in order to do so, given that this algorithm is able to capture some complex non-linearities that the basic LPM used could not detect. The first stage of this approach is studied in Figure A.2, that reports the variable importance measure described in Section 2 for each covariate. One can see that geography seems to matter in order to predict "long term unemployment" — defined in this particular setting as not having found a job six months after the experiment — as four department dummies appear to be among the top 10 of the variables used the most for splits in the forest built. Another interesting fact to underline is the importance of the various tightness indicators: most of them are among the top 50% of the variables used for splits, and those having the worse ranks — still not far from the median variable importance score — appear to be those for which there was other indicators highly correlated that were in competition for splits. Indeed, the tightness indicators that are ranked the worse are the ones based on the BMO survey, for which I included two versions: one based on a definition of markets at the department level, and one at the regional level. I find it interesting to notice this importance of tightness indicators in the prediction of outcome in the absence of treatment as the first use of these measure in Table 4 did not reveal (for most indicators) a significant correlation between them and the job finding probability. Given the importance of those variables in the construction of this forest, a candidate explanation might be that tightness — at least as measured by those indicators — interacts non-linearly with other covariates in order to predict the job finding probability.

I report in Table 6 the results of this procedure. They are rather surprising, as they suggest treatment effect heterogeneity of a magnitude that seems rather unbelievable: the treatment would decrease by around 6 percentage points the job finding probability of individuals with a low risk of LTU, and increase by about the same amount the job finding probability of individuals with a high risk of LTU. This seems far too large given the nature of the treatment, and suggests a risk of overfitting of the ML model built to predict the outcome in the absence of treatment. Indeed, if the model is subject to overfit, then it will associate negatively selected units to treated units predicted to have a high risk of LTU — therefore biasing upward the treatment effect estimated for individuals with a high predicted risk of LTU — and the reverse mechanism will be at play for individuals with a low predicted risk of LTU.

Given those concerns, I devote Section 6.1.2 to the elaboration of a robust analysis of treatment effect heterogeneity with respect to risk of LTU.

6.1.2 Robustness check: a word of caution regarding "honest" ML algorithms and their use in endogenous stratification analysis

The results of Table 6 suggesting some overfitting are rather surprising since the ML method I used, a `regression_forest` from the `grf` R package with a honesty parameter set to `TRUE`, is supposed to be robust to any risk of overfitting — I actually chose it because of this specificity, as I was aware of the concerns raised by Abadie et al. (2018) about an overfitting bias when

Table 6: Heterogeneous treatment effects with respect to risk of LTU
(ML measure)

	Dependent variable:	
	Job found after 6 months	
	(1)	(2)
Treatment	-0.065^{***}	-0.065^{***}
	(0.005)	(0.005)
Above median proba. of LTU — ML prediction	-0.292^{***}	-0.274^{***}
	(0.004)	(0.005)
Treatment × Above median proba. of LTU	0.128^{***}	0.128^{***}
	(0.005)	(0.005)
Constant	0.375^{***}	0.426^{***}
	(0.004)	(0.008)
Controls	No	Yes
Observations	149,286	149,286

*p<0.1; **p<0.05; ***p<0.01

Note: Robust clustered standard errors are reported in parentheses (clusters are the FAP code × Employment agency (ALE) level). These regressions are performed over the whole sample of job seekers (without missing values). The interaction term indicates whether or not there are significantly different treatment effects for individuals with an above (as compared to below) median probability/risk of long-term unemployment (LTU), as predicted by a "honest" regression forest. The significance test for the interaction term with the treatment variable (in bold) indicates whether or not there is treatment effect heterogeneity.

using endogenous stratification. Indeed, as it is supposed to be "honest", this function should be built such that any tree used to predict the outcome for a given unit is grown *without* using this unit. Facing those results, I decided to implement a first small test of this kind of model for overfitting. In order to do so, I built a regression forest model using exactly the same algorithm and tuning parameters on a random subsample of half the control units; I then computed the predicted probability of finding a job after 6 months — i.e. (1 - "risk of LTU") — for the controls units used to build the forest *and* for the other half of control units. I then consider that when the predicted probability is higher than 0.5, the prediction is that the individual found a job after 6 months. Finally, I compare these predictions with the actual outcomes, on the set of control units used to build the forest and on the set of control units that were left aside. The results are once again disturbing, as I do not find any sign of major overfitting by doing so: the proportion of right predictions is at 77.7% for control units used to build the forest, and 77.4% for control units left aside.

Yet, this kind of overfitting bias has already been documented when using causal forests to estimate heterogeneous treatment effects (Heller and Davis, 2017). I therefore still attempt to correct for such a bias. A way to get for sure an unbiased estimate would be to use the procedure described by Abadie et al. (2018), using a repeated split sample estimator, along the lines of the procedure implemented in Section 5.3. However, if I were to do so, I could not compute the associated standard error, since bootstrap repetitions would be far too computationally costly. Instead, I decide to "force" an honest estimation by creating myself one single partition of control and treated units, and build two ML prediction models based on the two different splits of control units. I then implement cross-fitting: I apply the predictions of the model built on one half of the control units to the other half of control units and one half of treatment units, and vice-versa. This way, I have for sure some honest estimation of the risk of LTU, that can be used to investigate treatment effect heterogeneity. I report the results in column (2) of Table 7. We do observe a significant change with the results presented in Table 6. They suggest that the treatment would decrease by around 1.5 percentage points the job finding probability of individuals with a low risk of LTU, and increase by about the same amount the job finding probability of individuals with a high risk of LTU. This is interesting as it still suggests that this treatment is more efficient for individuals who are the most in need of help. Notice that there are other examples in the literature suggesting that some job search assistance policies are more efficient for individuals with the worse employment prospect Altmann et al. (2018). On the contrary, a candidate explanation for the fact that the treatment decreases the job finding probability of individuals with a low risk of LTU is that it gives them the possibility to widen their employment research in search for a better match than the one they could find in the absence of LBB assistance.

Notice that one could think that the difference between results in column (2) of Table 7 and those of Table 6 are caused by the fact that the ML models used in Table 7 for honest estimation are built on only half of the sample of control units, while the ML model used in Table 6 was built using the whole sample of control units. It might be that this decrease in sample size decrease the quality of the models predicted, therefore changing the results obtained. In order to test whether those concerns are justified, I applied the predictions of the two models used for honest estimation to the samples of control units on which they were trained; in other words, I voluntarily break the honest estimation. I then run the same regression exploring treatment effect heterogeneity with respect to the risk of LTU — the results are presented in column (1) of Table 7. If the concerns expressed above were true, then the results obtained in this column should at least be different from the ones obtained in Table 6, and could even be similar to the ones of column (2) of Table 7. On the contrary, the estimates obtained are quite similar to the ones obtained in Table 6.

6.2 Causal tree(s)

In this section, I turn to the use of causal trees (Athey and Imbens, 2016a), presented in Section 2. Building a causal tree requires at first an honest splitting of the data: one split will be used in order to build the tree, and the other split will be used in order to estimate the conditional average treatment effects in each leaf. Indeed, this is crucial since trees are built by looking for heterogeneous treatment effects on the training split; as such, the estimations of CATEs on this same training split once the tree is built can be subject to an overfitting bias. I therefore divide my sample in two splits A and B. Given the fact that the literature has documented some inconsistent model selection when using the causal tree algorithm, I build two trees: one built/trained on A and estimated on B, and the other built on B and estimated on A. I present the resulting trees in Figures 3 and 4. In Tables A.6 and A.7, I report the results of regressions testing whether the CATEs computed for each leaf of the trees are indeed significantly different from zero. Those trees were built using the `causal_tree` function from `causalTree` R

Table 7: Honest estimation of treatment effect heterogeneity
with respect to risk of LTU

	Dependent variable:	
	Job found after 6 months	
	No cross-fitting	Cross fitting
	(1)	(2)
Treatment	-0.060^{***}	-0.016^{***}
	(0.003)	(0.003)
Above median proba. of LTU — ML prediction	-0.260^{***}	-0.174^{***}
	(0.003)	(0.003)
Treatment × Above median proba. of LTU — ML prediction	0.117^{***}	0.030^{***}
	(0.004)	(0.004)
Constant	0.357^{***}	0.313^{***}
	(0.002)	(0.002)
Observations	149,286	149,286

*p<0.1; **p<0.05; ***p<0.01

Note: The ML method used is the regression forest algorithm, with 2000 trees grown for each forests. Two models are built, using two different random half of control units. In column (1), I predict the risk of LTU for each half of control units (and two random half of treated units) using the model built on these control units (no cross-fitting, no honest estimation). In column (2), I use a honest estimation: the model used to predict risk of LTU for each half of control units has been built using the other half of control units.

package. I used tuning parameters such that each final leaf was supposed to contain at least 5% of the data: this was simply motivated by a will to limit the number of leafs of the final output, for the sake of interpretability. Other tuning parameters were set to default.

I indeed find two quite different trees depending on the split used as the training set. However, both trees share a quite interesting result: tightness indicators are massively used for splits. This suggests that tightness does interact with the treatment effect. Notice however that only few of the leafs of each trees have CATEs that are found significantly different from zero in the regressions of Tables A.6 and A.7. For the tree built on split A, it is leafs -0.0282, -0.0256, -0.0242 and 0.0259 (the names of the leafs correspond to the estimated CATE within each leaf, reported in the final cells of Figures 3 and 4). For the tree built on split B, it is leafs 0.0272 and 0.0208.

A last interesting observation to make is that the tightness measure indicating the number of difficult hirings has an important role in the tree built on split A — it is the variable used for the first split. It could be viewed as a further evidence — after the results of Table 4 that this variable might really be a dimension along which the treatment effect varies.

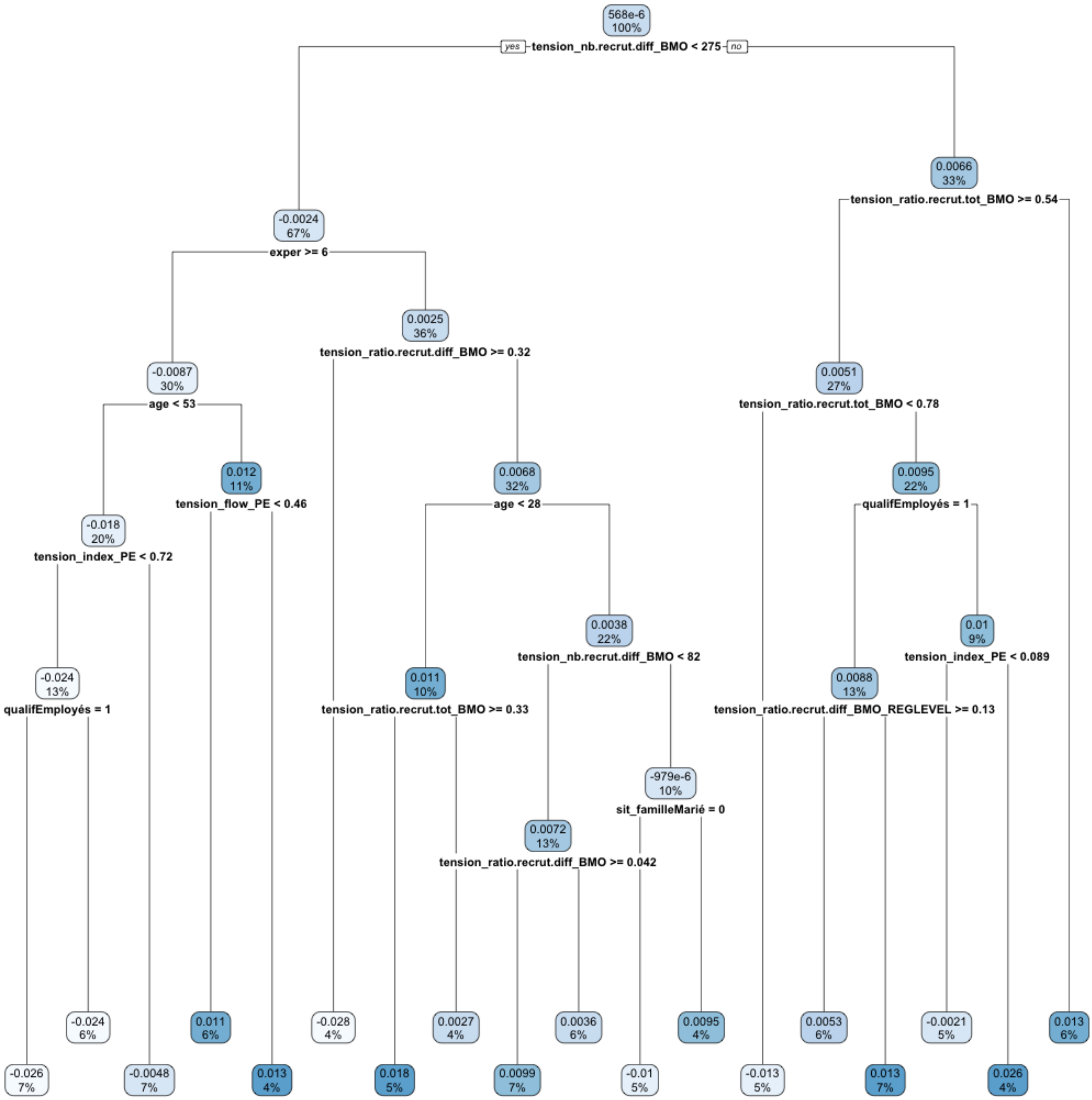


Figure 3: Causal tree built on split A of the data

Notes: Tree built using `causalTree` R package. At each node is indicated (i) the average treatment effect for individuals belonging to any leaf linked to this node, (ii) the share of total number of observations belonging to any leaf linked to this node, and (iii) the splitting variable used along with the level at which it is splitted. Each leaf (squares at the bottom of the tree) indicates (i) the average treatment effect for individuals belonging to this leaf and (ii) the share of total number of observations belonging to this leaf.

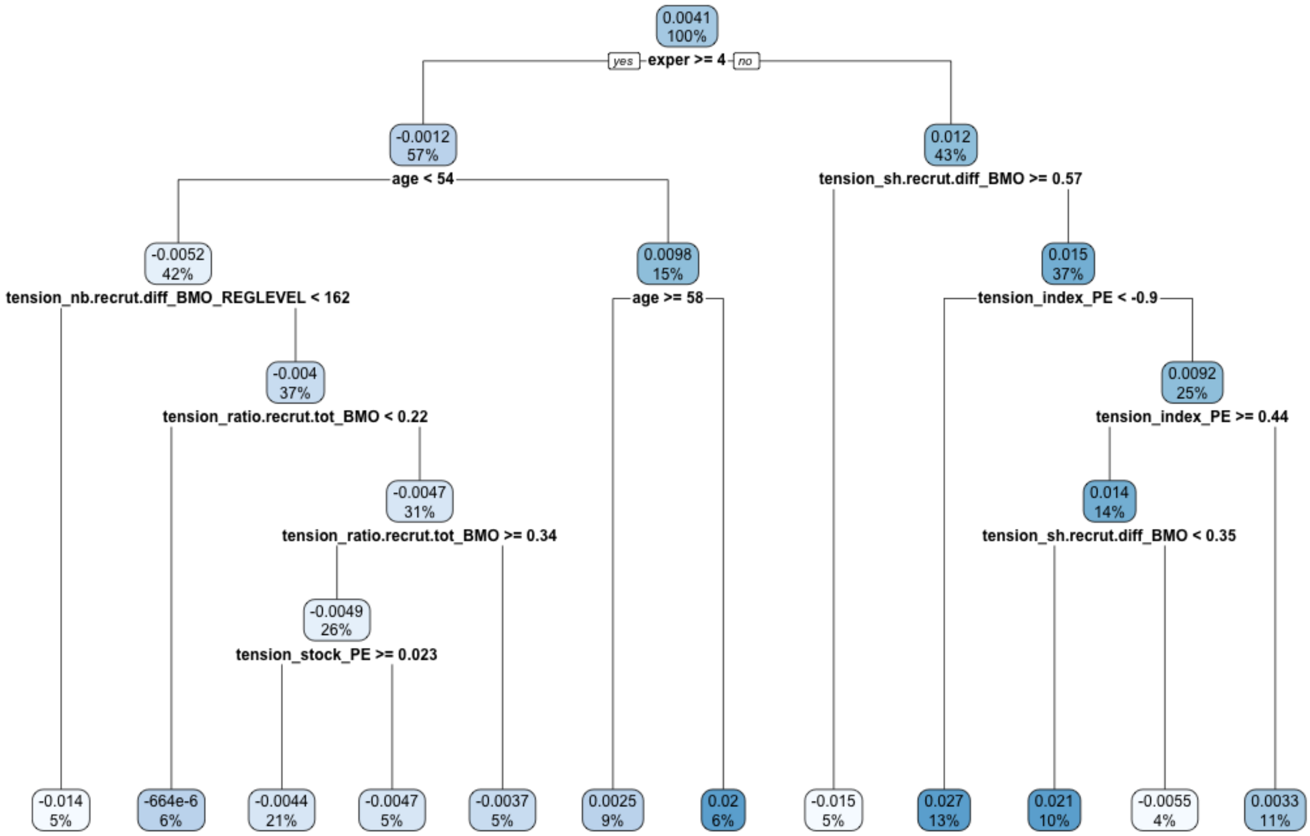


Figure 4: Causal tree built on split B of the data

Notes: Tree built using `causalTree` R package. At each node is indicated (i) the average treatment effect for individuals belonging to any leaf linked to this node, (ii) the share of total number of observations belonging to any leaf linked to this node, and (iii) the splitting variable used along with the level at which it is splitted. Each leaf (squares at the bottom of the tree) indicates (i) the average treatment effect for individuals belonging to this leaf and (ii) the share of total number of observations belonging to this leaf.

6.3 Building and exploring a Conditional Average Treatment Effect (CATE) function

I now turn to the use of a more flexible ML-estimated CATE function in order to explore treatment effect heterogeneity. I use the causal forest algorithm (Wager and Athey, 2017) described in Section 2.

6.3.1 Estimation of a CATE function

A CATE function is a function $\tau(x) = E\left[\tau_i | X_i = x\right]$, with $\tau_i = Y_i(1) - Y_i(0)$ the treatment effect for individual i. Causal trees allow to compute some simple CATE functions, that yield treatment effect estimates for groups of individual defined by interaction of covariates values. Causal forests, which are the result of combination of causal trees estimates, provide CATE functions that vary much more smoothly along covariates values; they can give a treatment effect estimate completely individualized (if enough trees are grown).

However, exploring the features of a CATE function built using a causal forest is far less straightforward than when using causal trees. I present in the paragraphs of the following section various attempts to do so, mostly based on propositions made by Susan Athey in her lectures available on AEA website[44].

The CATE function studied in the following section was built using the `causal_forest` function from `grf` R package. 2000 trees were grown, and honesty parameter was set to `TRUE`: I divided myself the sample between a training sample that would be used to grow the 2000 trees composing the forest, and a main sample that would be used for estimation of the CATEs. Other tuning parameters were set to default.

6.3.2 Exploring the CATE function features

Heatmaps Based on the suggestions of Susan Athey in her lectures, I built three heatmaps presented in Figures A.5, A.4, A.3. Each heatmap plots deciles of a tightness indicator against age deciles. Each cell color indicates the average value of the CATE function at a given intersection of the two variables plotted. The three indicators used are:

i. the ratio of the stock of vacancies over the stock of job seekers on each market, as recorded by PE in the STMT database (`tension_stock_PE`);

[44]See: `https://www.aeaweb.org/conference/cont-ed/2018-webcasts`

ii. the proportion of difficult hirings among total hirings scheduled on each market (`tension_ratio_recrut_tot_BM0`);

iii. the number of difficult hirings scheduled on each market (`tension_nb_recrut_diff_BM0`).

Surprisingly, all three heatmaps tend to suggest that there is more heterogeneity detected across age deciles than across deciles of the tightness measures. For `tension_stock_PE`, the heatmap tends to suggest the following: the higher this tightness indicator, the lower the treatment effect. This would go against the previous results obtained in Section 5. For the two other indicators, the heatmaps would suggest the reverse, in line with the results of 5. However, these heatmaps are not extremely informative at the end of the day, as there is not any information about the significativity of the difference between CATEs computed in each cell.

Covariates values across CATE quintiles Another suggestion of Susan Athey in her lectures is to study the covariates values across quintiles of predicted CATEs. For those familiar with this literature, or those having read Section 2 of this thesis, that might sound like a CLAN analysis. However, in contrast with the work of Chernozhukov et al. (2018a), Susan Athey does not provide any valid inference strategy to prove that predicted CATEs quintiles are relevant heterogeneity groups in the first place — which is what the analysis of GATEs allows to do in Chernozhukov et al. (2018a) approach (see Section 6.4 for an implementation of this methodology to this case). As such, even if we observe significant differences in covariates values across quintiles, nothing guarantees that there is significant treatment effect heterogeneity across those quintiles... That said, I still report the results of such an analysis in Figures A.6 (for various covariates) and A.7 (for the different tightness indicators). What can be noticed in Figure A.6 is an important variation of the educational level of individuals across quintiles of predicted CATEs. This is in line with what we were observing in the "classic" heterogeneity analysis in Table 3: the causal forest algorithm seems to detect, as in the classic analysis, potential treatment effect heterogeneity with respect to education. In Figure A.7, a striking fact is that while the average value of tightness indicators based on PE data seem to decrease across predicted CATEs quintiles, the reverse is true for indicators based on BMO survey. This is further evidence of the surprising decorrelation of these indicators.

Variable importance measure As used earlier in this work and presented in Section 2, a way to dig into the causal forest creating the CATE function is the variable importance measure, that is based on the number of times a variable was used to create a split. As such, it gives some extra (but indirect) information about the link between each variable and treatment effect heterogeneity. I report the variable importance measures for all covariates in Figure A.8. Once again, the importance of tightness indicators is striking: all of them are in the top 50% of covariates ranked by variable importance — despite the high correlation between some of them[45]. Notice as well that the tightness indicator based on the number of difficult hirings scheduled in each market is the first tightness indicator in terms of variable importance, and the seventh more important covariate: this is further evidence of the association of this variable with treatment effect heterogeneity.

A simpler feature of the CATE function: sorted individual treatment effects For this part, I was inspired by O'Neill and Weeks (2018), who just like me are using a CATE function built using a causal forest in order to study treatment effect heterogeneity in the context of a randomized experiment. The idea is to take advantage of the fact that the causal forest algorithm developed based on Wager and Athey (2017) allows for confidence interval estimation of each individual treatment effect estimates: it is therefore possible to plot every individual treatment effect estimates sorted by size, with a confidence interval; one can then observe whether there seem to be treatment effect heterogeneity revealed by the CATE function. I report the result in Figure 5. As one could have expected given the rather limited treatment effect heterogeneity uncovered so far, Figure 5 does not report an important treatment effect heterogeneity.
Notice however that this approach has its limitations. Indeed, as mentioned in Chernozhukov et al. (2018a), the point-wise inference developed by Wager and Athey (2017) — upon which confidence intervals plotted in Figure 5 are based — is valid only in low-dimensional settings (d $<$ log(n)) and with continuous variables, which is not the case here.

6.4 Best Linear Predictor (BLP), Sorted Group Average Treatment Effects (GATEs) and Classification Analysis (CLAN)

One last (and recent) way to approach the puzzle of heterogeneity exploration using ML is the one proposed by Chernozhukov et al. (2018a), and presented in Section 2. The main idea is to focus on "simple" features of the CATE function,so that a valid inference strategy could

[45]Indeed, for the tightness indicators based on the BMO survey, I included two versions: one based on a definition of markets at the department level, and one at the regional level. Those two versions are highly correlated by construction.

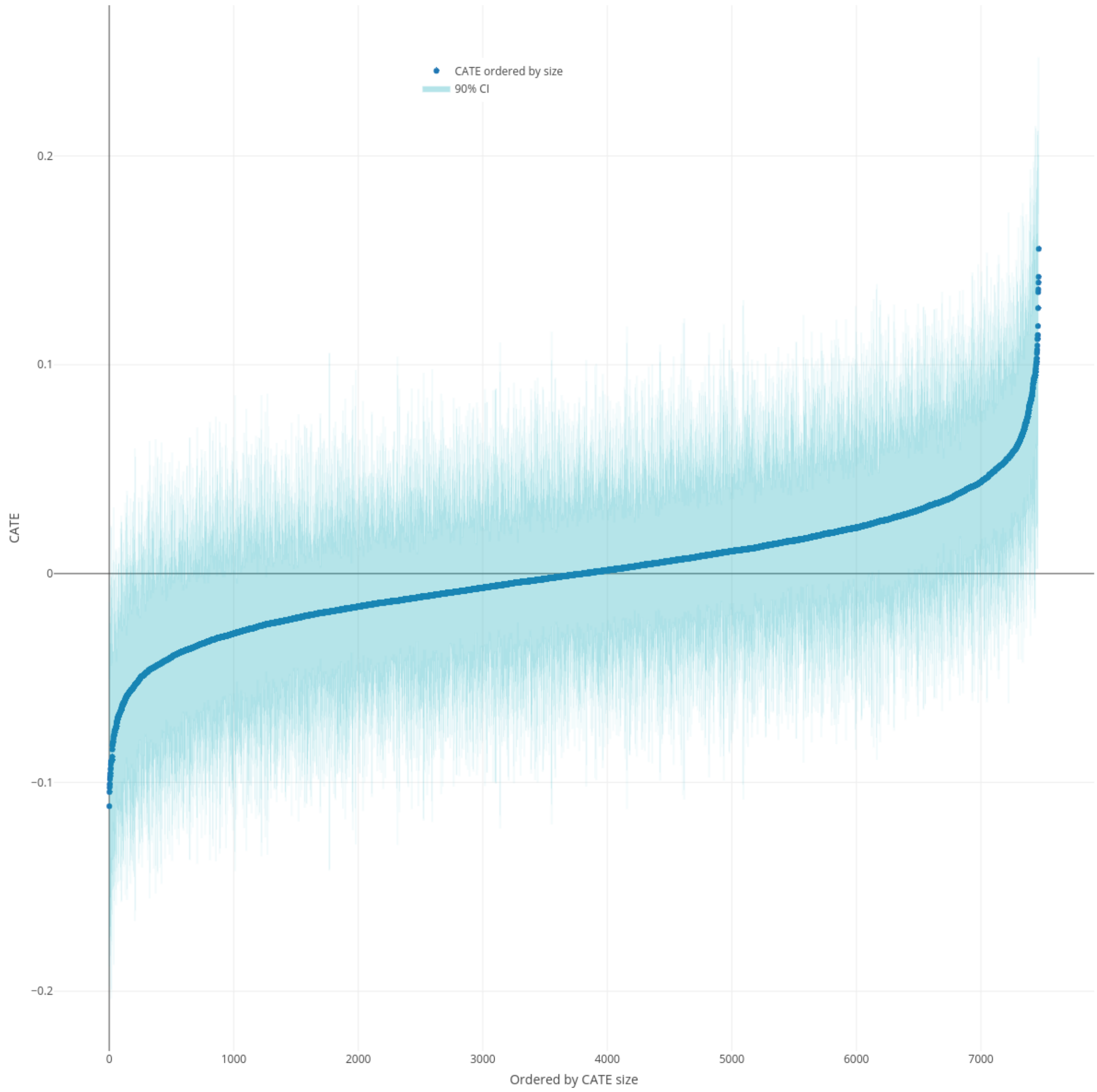


Figure 5: Estimate CATEs ordered by CATE size

Notes: In this figure, I present a plot of the estimated CATEs for each individuals, ordered by CATE size (see 6.3.2 for comments), along with the 90% confidence interval (in pale blue) for each estimate (using the standard errors estimate by the `causal_forest` function of `grf` R package. Since the plot function could not plot the 74,642 points (half of the whole data, corresponding to the random sample used for prediction), I plot a random sample of 10% of these observations (7,464 points).

be developed for those features. The authors propose to focus on three features (using their notations):

i. the Best Linear Predictor (BLP) of the true CATE function $s_0(Z)$ based on the ML proxy predictor $S(Z)$;

ii. the sorted Group Average Treatment Effects (GATEs), the average of conditional treatment effects $s_0(Z)$ by heterogeneity groups (quintiles of predicted CATEs) induced by the ML proxy predictor $S(Z)$;

iii. the CLassification ANalysis (CLAN) that reports the average characteristics of the most and least affected units — as defined by the ML proxy predictor $S(Z)$.

I implement the strategy described by the authors in order to identify those three main features of a CATE function built using causal forest. Notice that any other ML method could be used in order to estimate a CATE function before identifying those three features (elastic net, neural networks, boosted trees etc.). In Chernozhukov et al. (2018a), the authors indicate that they find the random (causal) forest and elastic net to be the most efficient ML methods. Since I have been using tree-based methods all along this work, I continue with a causal forest algorithm. I include in the algorithm all original covariates contained in my data set (sociodemographic characteristics, search process information, tightness indicators from PE STMT database and BMO survey). I do not include a prediction of LTU risk nor an estimate of market job finding rate as the goal is to determine whether ML allows to uncover, without extra search on the data, some treatment effect heterogeneity.

Best Linear Predictor (BLP) The first main feature proposed by Chernozhukov et al. (2018a) is the BLP of treatment effects. One key element of this BLP is that it is composed of two coefficients: the first (β_1 in the authors' notations) stands for the ATE estimate, while the other (β_2 in the authors' notations), that multiplies the ML prowy predictor for the CATE, is the "heterogeneity loading parameter", which indicates whether there is indeed some heterogeneity with respect to the ML prediction for the CATE. The BLP can eventually be written as follows (using authors' notations):

$$\text{BLP}\left[s_0(Z)|S(Z)\right] = \beta_1 + \beta_2(S(Z) - \text{E}S)$$

with $s_0(Z)$ the true CATE function, $S(Z)$ the ML proxy predictor for the true CATE, and ES the average of $S(Z)$. Following the strategy exposed by the authors in their paper, I identify β_1 and β_2. The results are presented in Table 8. The main advantage of BLP estimation is that it offers a test for treatment effect heterogeneity: if β_2 is found to be significantly different from zero, it means that there is significant treatment effect heterogeneity.

Table 8: Coefficients of the Best Linear Predictor (BLP)

Coefficient	Point estimate	L.B. (90% CI)	U.B. (90% CI)
β_1 (ATE estimate)	0.0036	-0.0022	0.0094
β_2 (Heterogeneity loading parameter)	0.0617	-0.1585	0.2819

Note: Medians over 100 different splits of the data. The ML procedure used to predict treatment effect heterogeneity is a causal forest, from the `grf` R package. 2000 trees were grown for each forest used. Lower and upper bounds reported for a 90% confidence interval (medians over 100 splits).

However, what we observe in Table 8 is that β_2 is not found to be significantly different from 0. Notice that this does not constitute a proof of the absence of treatment effect heterogeneity; it simply indicates that the ML method used was not able to provide a proxy predictor for treatment effect heterogeneity that was relevant.

Sorted Group Average Treatment Effects (GATEs) The GATEs analysis consist in identifying the average of conditional treatment effects $s_0(Z)$ by heterogeneity groups (quintiles of predicted CATEs G1, G2, ..., G5) induced by the ML proxy predictor $S(Z)$. If the ML proxy predictor is relevant, then it can be that the average treatment effects across quintiles of $S(Z)$ differ. The basic test is: $\mathrm{E}\left[s_0(Z)|G_1\right] = \ldots = \mathrm{E}\left[s_0(Z)|G_K\right]$.

I report the result of the estimation of GATEs using the strategy proposed by Chernozhukov et al. (2018a) in Table 9.

Given the results of Table 8 for the BLP, it is not that surprising to discover in Table 9 that we cannot reject the null hypothesis of equality of GATEs: $\mathrm{E}\left[s_0(Z)|G_1\right] = \ldots = \mathrm{E}\left[s_0(Z)|G_K\right]$. This means that our ML proxy predictor of the true CATE does not provide relevant heterogeneity groups to analyse.

CLassification ANalysis (CLAN) The CLAN consists in reporting the average characteristics of the most and least affected units — as defined by the ML proxy predictor $S(Z)$ quintiles. The "issue" here is the following: given the results presented in Table 9, we already know that

Table 9: Coefficients for the sorted Grouped Average Treatment Effects (GATEs)

Coefficient	Point estimate	L.B. (90% CI)	U.B. (90% CI)
γ_1	0.0060	−0.0070	0.0190
γ_2	−0.0038	−0.0169	0.0092
γ_3	0.0029	−0.0101	0.0160
γ_4	0.0002	−0.0128	0.0132
γ_5	0.0128	−0.0002	0.0258

Note: Medians over 100 different splits of the data. The ML procedure used to predict treatment effect heterogeneity is a causal forest, from the `grf` R package. 2000 trees were grown for each forest used. Lower and upper bounds reported for a 90% confidence interval (medians over 100 splits).

there is no significant differences in the GATEs. As such, the CLAN is not truly informative, as it gives us information about the characteristics of groups that are not found to be significantly differently affected by the treatment. I still report the results of this analysis in Table 10. Notice that the issue with the CLAN approach is that at this stage, we are once again facing the issue of MHT: how to account for the fact that we are testing for significant differences in covariates averages for mutliple covariates? Moreover, it only provides a superficial understanding of the mechanisms driving heterogeneity, as we do not test for a different in the value of interacted covariates for instance.

Table 10: CLassification ANalysis (CLAN)

Variable name	Avr. G1	G1 LB 90%	G1 UB 90%	Avr. G5	G5 LB 90%	G5 UB 90%	G1≠G5
Sociodemographic variables							
age	36.1306	35.5042	36.7571	40.1222	39.5084	40.7361	∗
contrat_rechCDI	0.9351	0.9211	0.9491	0.9229	0.9089	0.9369	.
contrat_rechSaisonnier	0.0040	0.0027	0.0054	0.0117	0.0104	0.0131	∗
depart_nb49	0.0384	0.0353	0.0415	0.0297	0.0266	0.0328	∗
depart_nb53	0.0090	0.0075	0.0106	0.0064	0.0048	0.0079	.
depart_nb54	0.0408	0.0375	0.0441	0.0397	0.0364	0.0430	.
depart_nb55	0.0067	0.0053	0.0081	0.0050	0.0036	0.0065	.
depart_nb57	0.0683	0.0644	0.0723	0.0526	0.0486	0.0566	∗
depart_nb72	0.0265	0.0240	0.0291	0.0212	0.0187	0.0238	∗
depart_nb75	0.1395	0.1333	0.1456	0.1882	0.1821	0.1943	∗
depart_nb77	0.0675	0.0634	0.0716	0.0626	0.0585	0.0668	.
depart_nb78	0.0782	0.0740	0.0824	0.0562	0.0520	0.0604	∗
depart_nb85	0.0406	0.0378	0.0434	0.0316	0.0288	0.0344	∗
depart_nb88	0.0178	0.0154	0.0203	0.0208	0.0183	0.0232	.
depart_nb91	0.0707	0.0668	0.0745	0.0526	0.0487	0.0565	∗
depart_nb92	0.0747	0.0699	0.0795	0.1106	0.1058	0.1154	∗
depart_nb93	0.0838	0.0786	0.0890	0.1253	0.1202	0.1305	∗
depart_nb94	0.0800	0.0757	0.0844	0.0608	0.0564	0.0652	∗
depart_nb95	0.0681	0.0639	0.0724	0.0679	0.0636	0.0721	.
educBAC_2	0.1554	0.1494	0.1614	0.1440	0.1380	0.1501	.
educBAC_3_4	0.1060	0.1008	0.1113	0.1141	0.1089	0.1194	.
educBAC_5_plus	0.1220	0.1163	0.1277	0.1746	0.1689	0.1802	∗
educBelow_BAC	0.3468	0.3375	0.3560	0.3177	0.3083	0.3270	∗
educNo_educ	0.0140	0.0113	0.0167	0.0279	0.0252	0.0306	∗
exper	7.3226	7.1455	7.4996	6.3179	6.1388	6.4969	∗
female	0.4609	0.4502	0.4716	0.4952	0.4846	0.5059	∗
motifDémission	0.0196	0.0175	0.0217	0.0143	0.0122	0.0164	∗
motifFin_de_contrat	0.2365	0.2289	0.2440	0.2657	0.2582	0.2732	∗
motifFin_de_maternité_maladie	0.0367	0.0332	0.0402	0.0480	0.0445	0.0514	∗
motifLicenciement	0.1888	0.1822	0.1953	0.1305	0.1239	0.1371	∗
motifLicenciement_éco	0.0380	0.0348	0.0412	0.0350	0.0317	0.0382	.
nationamnord	0.0008	0.0002	0.0014	0.0015	0.0009	0.0021	.
nationamsud	0.0047	0.0036	0.0058	0.0037	0.0026	0.0048	.
nationasie	0.0092	0.0074	0.0109	0.0132	0.0114	0.0149	∗
nationeurope	0.0373	0.0339	0.0407	0.0368	0.0333	0.0402	.
nationfrench	0.8877	0.8744	0.9011	0.8589	0.8455	0.8724	∗
nenf	0.7610	0.7385	0.7835	0.7513	0.7287	0.7738	.
projentr	0.0980	0.0929	0.1031	0.0943	0.0892	0.0994	.
qualifEmployés	0.6302	0.6181	0.6423	0.6623	0.6503	0.6743	∗
qualifOuvriers	0.1230	0.1177	0.1284	0.0841	0.0788	0.0895	∗
qualifTechniciens	0.0944	0.0897	0.0992	0.0833	0.0785	0.0881	∗
regionLorraine	0.1336	0.1278	0.1394	0.1181	0.1123	0.1239	∗
regionPays_de_la_Loire	0.2039	0.1974	0.2104	0.1577	0.1511	0.1642	∗
sit_familleDivorcé	0.0804	0.0752	0.0857	0.1015	0.0962	0.1067	∗
sit_familleMarié	0.3657	0.3556	0.3758	0.4112	0.4012	0.4212	∗
sit_familleVeuf	0.0053	0.0037	0.0068	0.0084	0.0068	0.0099	.
temps_rechTemps_partiel	0.0882	0.0827	0.0936	0.1172	0.1118	0.1227	∗
zus	0.0746	0.0703	0.0789	0.0658	0.0615	0.0702	∗
Labor market tightness information							
tension_flow_PE	0.4883	0.4792	0.4973	0.4808	0.4717	0.4898	.
tension_index_PE	0.1612	0.1214	0.2011	-0.2096	-0.2495	-0.1698	∗
tension_nb.recrut.diff_BMO	270.1134	260.5204	279.7065	462.1137	452.9051	471.3222	∗
tension_nb.recrut.diff_BMO_REGLEVEL	1,478.6270	1,440.2590	1,516.9960	2,215.6040	2,179.1060	2,252.1030	∗
tension_ratio.recrut.diff_BMO	0.2453	0.2370	0.2535	0.3284	0.3204	0.3365	∗
tension_ratio.recrut.diff_BMO_REGLEVEL	0.1246	0.1216	0.1276	0.1370	0.1340	0.1400	∗
tension_ratio.recrut.tot_BMO	0.7087	0.6900	0.7275	0.9291	0.9109	0.9474	∗
tension_ratio.recrut.tot_BMO_REGLEVEL	0.3538	0.3466	0.3610	0.4027	0.3956	0.4098	∗
tension_sh.recrut.diff_BMO	0.3566	0.3510	0.3623	0.3123	0.3065	0.3181	∗
tension_stock_PE	0.0533	0.0523	0.0542	0.0441	0.0431	0.0451	∗

Note: Medians over 100 different splits of the data. The ML procedure used to predict treatment effect heterogeneity is a causal forest, from the `grf` R package. 2000 trees were grown for each forest used. Columns (1) and (4) report average value of each covariate (median over 100 splits). Columns (2)-(3) and (5)-(6) report confidence intervals (each bound is the median of lower/upper bound over 100 splits). Column (7), the last column, is filled with an (*) when the confidence intervals for the average value in G1 and G5 do not overlap: this is a conservative test for equality between those two averages — i.e. it can be that column (7) does not report a (*) and that the two averages are still significantly different, since confidence intervals overlapping is not enough to fail to reject equality between two means. I am forced to use this test since I did not implemented the formula from Chernozhukov et al. (2018b) to compute adjusted p-values.

7 Conclusion

In this thesis, I study the effect of an online job search assistance program provided by *La Bonne Boîte* (LBB), a French start-up associated with Pôle emploi (PE). I take advantage of a previous experiment made by the French public employment services (Pôle emploi, PE henceforth) in 2015, which provides some exogenous variation in the use of this platform. I focus on the heterogeneity analysis of this treatment, using two main different approaches.

The first approach builds upon the theoretical background of the unemployment equilibrium theory, and therefore focuses on the analysis of the treatment effect heterogeneity with respect to various different labour market tightness indicators collected in different data sources. If anything, two main results stand out. (i) Quite surprisingly, the different tightness indicators studied are highly decorrelated, making it difficult to corroborate the rare significant results obtained when studying in a "classic" way heterogeneous treatment effects. (ii) The set of significant results obtained suggest that the treatment effect is *increasing* in labour market tightness. At least two simple models of the treatment are consistent with such results. One is the simple "threat effect" model: the e-mail of PE leads to an increase in job seekers' search intensity, which has a higher effect on markets in high tightness since the matching rate between applications and vacancies is higher on these markets. A second one is a model in which online job search assistance is considered as a *vacancy sorter*: it increases the probability that a match between a vacancy and an application actually results in a recruitment, which is similar to an increase in the aggregate matching efficiency; this has more impact when tightness is high as the matching rate between applications and vacancies is higher on these markets. I also document some evidence of a larger treatment effect for individuals with worser employment prospects. This is in line with other empirical evidence in the literature evaluating job search assistance programs Altmann et al. (2018).

In a second phase, I adopt a more data-driven approach, with the help of the new machine learning (ML) techniques developed for heterogeneity analysis. I focus on tree-based techniques and forests, which have been central in the development of these techniques and have been reported to be efficient in some of these works (Chernozhukov et al., 2018a). In the future, I could still widen my analysis to other ML techniques, especially elastic nets, that have also been the subject of several papers in this literature. The results of this analysis underline some limitations of ML techniques to uncover treatment effect heterogeneity, especially as the main ML-specific test for treatment effect heterogeneity developed by Chernozhukov et al. (2018a) concludes that ML is unable to detect any heterogeneity. Yet, this might be not that surprising after all given the lack of statistical power (low take-up) and the probably low order of magnitude of the treatment effect studied. Still, I provide applications of a large part of the existing ML techniques for treatment effect heterogeneity, trying to take advantage of each of them to document which are the dimensions that are likely to be important to study treatment effect heterogeneity in my setting. I hope that this part of the work presented here will humbly contribute to the growing literature trying to establish the qualities and limitations of these new ML techniques in applied econometrics.

References

Abadie, A., Chingos, M. M., and West, M. R. (2018). Endogenous Stratification in Randomized Experiments. *The Review of Economics and Statistics*, C(4):567–580.

Altmann, S., Falk, A., Jäger, S., and Zimmermann, F. (2018). Learning about job search: A field experiment with job seekers in Germany. *Journal of Public Economics*, 164:33–49.

Athey, S. and Imbens, G. (2016a). Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences.*

Athey, S. and Imbens, G. (2016b). The Econometrics of Randomized Experiments. In Banerjee, A. V. and Duflo, E., editors, *Handbook of Economic Field Experiments*, number July, pages 73–140.

Athey, S. and Imbens, G. W. (2017). The State of Applied Econometrics: Causality and Policy Evaluation. *Journal of Economic Perspectives*, 31(2):3–32.

Athey, S. and Imbens, G. W. (2019). Machine Learning Methods Economists Should Know About.

Athey, S., Tibshirani, J., and Wager, S. (2018). Generalized random forests. *Annals of Statistics (forthcoming).*

Athey, S. and Wager, S. (2018). Efficient Policy Learning.

Autor, D. H. (2001). Wiring the Labor Market. *Journal of Economic Perspectives*, 15(1):25–40.

Behaghel, L., Crépon, B., and Gurgand, M. (2014). Private and Public Provision of Counseling to Job Seekers: Evidence from a Large Controlled Experiment. *American Economic Journal: Applied Economics*, 6(4):142–174.

Belloni, A., Chernozhukov, V., and Hansen, C. (2014). High-Dimensional Methods and Inference on Structural and Treatment Effects. *Journal of Economic Perspectives*, 28(2):1–23.

Belot, M., Kircher, P., and Muller, P. (2018a). How wage announcements affect job search – a field experiment.

Belot, M., Kircher, P., and Muller, P. (2018b). Providing Advice to Job Seekers at Low Cost : An Experimental Study on On-Line Advice. *(forthcoming) Review of Economic Studies.*

Bertrand, M., Crépon, B., Marguerie, A., and Premand, P. (2017). Contemporaneous and Post-Program Impacts of a Public Works Program: Evidence from Côte d'Ivoire.

Breiman, L. (2001). Statistical Modeling : The Two Cultures. *Statistical Science*, 16(3):199–215.

Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J. (1984). *Classification and regression trees.*

Bren**ci**c, V. (2014). Search Online: Evidence from Information Acquisition on Online Job Boards and Resume Banks. *Journal of Economic Psychology*, 42:112–125.

Brunnermeier, M. K. and Parker, J. A. (2005). Optimal Expectations. *American Economic Review*, 95(4):1092–1118.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., and Robins, J. (2018a). Double / debiased machine learning for treatment and structural parameters. 21:C1–C68.

Chernozhukov, V., Demirer, M., Duflo, E., and Fernandez-Val, I. (2018b). Generic machine learning inference on heterogenous treatment effects in randomized experiments.

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

Duflo, E. (2018). Machinistas meet randomistas : useful ML tools for empirical researchers.

Duflo, E., Kremer, M., and Robinson, J. (2011). Nudging farmers to use fertilizer: Theory and experimental evidence from kenya. *American Economic Review*, 101(6):2350–90.

Faberman, J. R. and Kudlyak, M. (2013). The Intensity of Job Search and Search Duration.

Faberman, J. R. and Kudlyak, M. (2016). What does online job search tell us about the labor market? *Economic Perspectives*, pages 1–15.

Farcomeni, A. (2007). A review of modern multiple hypothesis testing, with particular attention to the false discovery proportion. *Statistical Methods in Medical Research*, 17:347–388.

Heller, S. B. and Davis, J. M. (2017). Using Causal Forests to Predict Treatment Heterogeneity: An Application to Summer Jobs. *American Economic Review: Papers & Proceedings*, 107(5):546–550.

Horton, J. J. (2017). The Effects of Algorithmic Labor Market Recommendations : Evidence from a Field Experiment. *Journal of Labor Economics*, 35(2).

Knaus, M. C., Lechner, M., and Strittmatter, A. (2018). Heterogeneous Employment Effects of Job Search Programmes : A Machine Learning Approach.

Koszegi, B. (2006). Ego Utility , Overconfidence , and Task Choice. *Journal of the European Economic Association*, 4(4):673–707.

Kroft, K. and Pope, D. G. (2014). Does Online Search Crowd Out Traditional Search and Improve Matching Efficiency ? Evidence from Craigslist. *Journal of Labor Economics*, 32(2):259–303.

Kuhn, P. J. (2014). The internet as a labor market matchmaker: how effective are online methods of worker recruitment and job? *IZA World of Labor*, (May):1–10.

Kuhn, P. J. and Mansour, H. (2013). Is internet job search still ineffective? *The Economic Journal*, 124:1213–1233.

Kuhn, P. J. and Skuterud, M. (2004). Internet Job Search and Unemployment Durations. *American Economic Review*, 94(1):212–232.

List, J. A., Shaikh, A. M., and Xu, Y. (2016). Multiple Hypothesis Testing in Experimental Economics.

Mueller, A. I., Spinnewijn, J., and Topa, G. (2019). Job Seekers ' Perceptions and Employment Prospects : Heterogeneity , Duration Dependence and Bias .

Mullainathan, S. and Spiess, J. (2017). Machine Learning: An Applied Econometric Approach. *Journal of Economic Perspectives*, 31(2):87–106.

Nekoei, A. and Weber, A. (2017). Does Extending Unemployment Benefits Improve Job Quality? *American Economic Review*, 107(2)(9034):527–561.

O'Neill, E. and Weeks, M. (2018). Causal Tree Estimation of Heterogeneous Household Response to Time-Of-Use Electricity Pricing Schemes.

Orand, M. (2018). Comment mesurer les tensions sur le marché du travail en France ? Technical report, DARES.

Petrongolo, B. and Pissarides, C. A. (2001). Looking into the Black Box : A Survey o the Matching Function. 39(2):390–431.

Pissarides, C. A. (2000). *Equilibrium Unemployment Theory*. The MIT Press.

Rogerson, R., Shimer, R., and Wright, R. (2006). Search-Theoretic Models of the Labor Market : A Survey. *Journal of Economic Literature*, 43(4):959–988.

Shimer, R. (2004). Search Intensity.

Signorelli, S. (2019). Do Skilled Migrants Compete with Native Workers ? Analysis of a Selective Immigration Policy.

Spinnewijn, J. (2015). Unemployed but optimistic: Optimal insurance design with biased beliefs. *Journal of the European Economic Association*, 13(1):130–167.

Strittmatter, A. (2019). What Is the Value Added by Using Causal Machine Learning Methods in a Welfare Experiment Evaluation ?

Varian, H. R. (2010). Computer Mediated Transactions (Richard T. Ely Lecture). *American Economic Review: Papers & Proceedings*, 100(May):1–10.

Varian, H. R. (2014). Big Data : New Tricks for Econometrics. *Jou*, 28(2):3–28.

Wager, S. and Athey, S. (2017). Estimation and Inference of Heterogeneous Treatment Effects using Random Forests.

Zeileis, A., Hothorn, T., and Hornik, K. (2008). Model-based Recursive Partitioning. *Journal of Computational and Graphical Statistics*, 17:492–514.

Appendices

A Tables and Graphs

Table A.1: LATE of following LBB link, at different time horizons

	Dependent variable:					
	Job found after 6 m.		Job found after 12 m.		Job found after 24 m.	
	(1)	(2)	(3)	(4)	(5)	(6)
Followed LBB link	0.012	0.012	0.003	0.003	0.010	0.010
	(0.012)	(0.011)	(0.013)	(0.013)	(0.014)	(0.013)
Constant	0.226***	0.518***	0.342***	0.732***	0.460***	0.926***
	(0.002)	(0.006)	(0.002)	(0.007)	(0.002)	(0.007)
Controls	No	Yes	No	Yes	No	Yes
Observations	149,311	149,311	149,311	149,311	149,311	149,311

*p<0.1; **p<0.05; ***p<0.01

Note: Robust clustered standard errors are reported in parentheses (clusters are the FAP code × Employment agency (ALE) level). These regressions are performed over the whole sample of job seekers (without missing values). Controls include age, sex, level of education, experience in the job for which the job seeker if searching, and qualification (categorical variable indicating whether the individual is a blue collar, clerical worker etc.). The dependent variables are dummies indicating whether the individual has found a job x months after the beginning of the experiment (i.e. December 2015). The instrument in our 2SLS regression is the assignment variable to treated or control groups (only the treated group received an e-mail directing towards LBB websited).

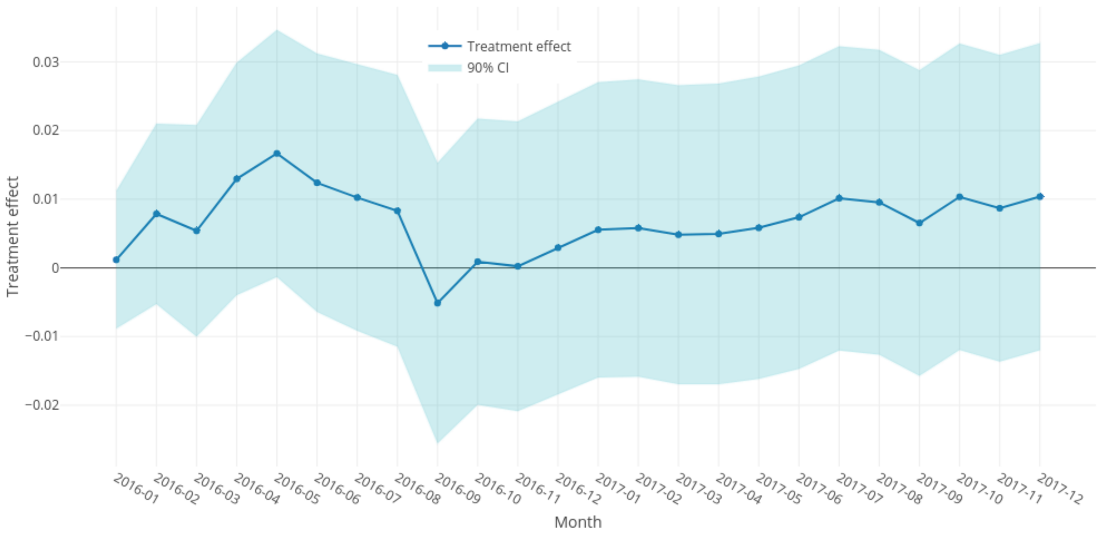


Figure A.1: Timing of treatment effect: LATE of following LBB link at different time horizons

Notes: In this figure, we report the LATE of following LBB link for different time horizons, which amounts to repeating the regression presented in Table A.1 for all available dependent variables, from month = 1 to month = 24. The LATE for month "2017-01" is the effect of the following LBB link on the probability of having found a job between January 2016 and January 2017 (month = 12 in this case), as estimated by a 2SLS regression.

Table A.2

	Dependent variable:					
	Job found after 6 months					
	Stock ratio (PE)	Flow ratio (PE)	Index (PE)	Nb. diff. hirings (BMO)	Stock ratio (BMO)	Sh. diff. hirings (BMO)
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	0.005	0.002	0.002	−0.001	0.001	0.008*
	(0.003)	(0.004)	(0.002)	(0.003)	(0.003)	(0.004)
Tightness measure	0.170	−0.010	0.003		−0.004	0.061***
	(0.137)	(0.016)	(0.003)		(0.004)	(0.023)
Treatment × Tightness measure	−0.045	0.001	0.0003		0.002	−0.017
	(0.031)	(0.004)	(0.001)		(0.003)	(0.013)
Constant	0.508***	0.521***	0.515***	0.535***	0.520***	0.494***
	(0.019)	(0.017)	(0.017)	(0.016)	(0.017)	(0.019)
Observations	149,286	149,286	149,286	149,286	149,286	149,286

*p<0.1; **p<0.05; ***p<0.01

Note: Robust clustered standard errors are reported in parentheses (clusters are the FAP code × Employment agency (ALE) level). These regressions are performed over the whole sample of job seekers (without missing values). Controls include age, sex, level of education, experience in the job for which the job seeker if searching, and qualification (categorical variable indicating whether the individual is a blue collar, clerical worker etc.). The dependent variable is a dummy indicating whether the individual has found a job 6 months after the beginning of the experiment (i.e. December 2015). The name of each column indicates the tightness measure used in the regression: the corresponding variable names are, in columns order, `"tension_stock_PE"`, `"tension_flow_PE"`, `"tension_index_PE"`, `"tension_nb.recrut.diff_BMO"`, `"tension_ratio.recrut.tot_BMO"`, `"tension_sh.recrut.diff_BMO"`. The significance test for the interaction term with the treatment variable (in bold) indicates whether or not there is treatment effect heterogeneity. In column (4), I included as control a proxy for the size of the market, namely total the number of hirings scheduled on the market (as measured by BMO), as the tightness indicator as hand in column (4) is not re-scaled to take this into account. Including this control in any of the 5 other regressions presented in this table was not changing the significance of the interaction term coefficient.

Table A.3: Heterogeneous treatment effects w.r.t. market job finding rate

	Dependent variable:			
	Job found after 6 months			
	Market job finding rate		Market job finding rate (residualized)	
	(1)	(2)	(3)	(4)
Treatment	0.027***	0.025***	0.028***	0.027***
	(0.004)	(0.003)	(0.004)	(0.003)
Above median market job finding rate	0.108***	0.074***		
	(0.007)	(0.006)		
Treatment	−0.050***	−0.047***		
× Above median market job finding rate	(0.006)	(0.006)		
Above median market job finding rate (residualized)			0.074***	0.076***
			(0.006)	(0.005)
Treatment			−0.051***	−0.050***
× Above median market job finding rate (residualized)			(0.006)	(0.006)
Constant	0.173***	0.410***	0.190***	0.410***
	(0.004)	(0.012)	(0.007)	(0.012)
Controls	No	Yes	No	Yes
Observations	149,286	149,286	149,286	149,286

*p<0.1; **p<0.05; ***p<0.01

Note: Robust clustered standard errors are reported in parentheses (clusters are the FAP code × Employment agency (ALE) level). These regressions are performed over the whole sample of job seekers (without missing values). Controls include region (3 levels) and aggregated FAP codes (87 levels) fixed effects, age, sex, level of education, experience in the job for which the job seeker if searching, and qualification (categorical variable indicating whether the individual is a blue collar, clerical worker etc.). The dependent variable is a dummy indicating whether the individual has found a job 6 months after the beginning of the experiment (i.e. December 2015). For the first two regressions (first two columns), the variable used as an "average job finding probability" at the market level is simply the average job finding rate after 6 months among control individuals in a given market (as defined by department (17 levels) × aggregated FAP codes (87 levels). In the two last regressions (two last columns), I use the residual of the average job finding rate as predicted by region (3 levels) and aggregated FAP codes (87 levels) fixed effects, in order to use only variation in average job finding probability across department within a same region, and within a same aggregated FAP code. The significance test for the interaction term with the treatment variable (in bold) indicates whether or not there is treatment effect heterogeneity.

Table A.4: LPM for Proba. of Long-Term Unemployment

Dependent variable: Job found after 6 months	
depart_nb49	−0.017*
	(0.010)
depart_nb53	0.002
	(0.016)
depart_nb54	−0.054***
	(0.009)
depart_nb55	−0.045***
	(0.017)
depart_nb57	−0.053***
	(0.008)
depart_nb72	−0.039***
	(0.011)
depart_nb75	−0.057***
	(0.007)
depart_nb77	−0.029***
	(0.008)
depart_nb78	−0.023***
	(0.008)
depart_nb85	0.050***
	(0.010)
depart_nb88	−0.068***
	(0.012)
depart_nb91	−0.013
	(0.008)
depart_nb92	−0.045***
	(0.008)
depart_nb93	−0.063***
	(0.007)
depart_nb94	−0.051***
	(0.008)
depart_nb95	−0.051***
	(0.008)
zus	0.002
	(0.006)
age	−0.007***
	(0.0002)
exper	0.002***
	(0.0002)
qualifEmployés	−0.022***
	(0.005)
qualifOuvriers	−0.014*
	(0.007)
qualifTechniciens	−0.008
	(0.007)
sit_familleDivorcé	0.024***
	(0.005)
sit_familleMarié	0.031***

(To be continued)

Dependent variable: Job found after 6 months	
	(0.004)
sit_familleVeuf	0.024
	(0.016)
nationamnord	0.008
	(0.041)
nationamsud	0.003
	(0.023)
nationasie	−0.051***
	(0.015)
nationeurope	−0.004
	(0.009)
nationfrench	−0.006
	(0.006)
contrat_rechCDI	0.016**
	(0.006)
contrat_rechSaisonnier	0.084***
	(0.019)
temps_rechTemps_partiel	−0.097***
	(0.005)
female	−0.007**
	(0.003)
tension_stock_PE	0.119***
	(0.036)
tension_nb.recrut.diff_BMO	−0.00001**
	(0.00000)
educBAC_2	0.027***
	(0.005)
educBAC_3_4	0.008
	(0.006)
educBAC_5_plus	0.031***
	(0.006)
educBelow_BAC	−0.012***
	(0.004)
educNo_educ	−0.031***
	(0.010)
Constant	0.527***
	(0.013)
Observations	74,547
Note:	*p<0.1; **p<0.05; ***p<0.01

Table A.5: Heterogeneous treatment effects with respect to risk of LTU
(OLS measure)

	Dependent variable:	
	Job found after 6 months	
	(1)	(2)
Treatment	0.001	0.001
	(0.003)	(0.003)
Above median proba. of LTU — OLS prediction	-0.148^{***}	-0.064^{***}
	(0.007)	(0.005)
Treatment × Above median proba. of LTU	0.002	0.002
	(0.003)	(0.003)
Constant	0.300^{***}	0.463^{***}
	(0.007)	(0.016)
Controls	No	Yes
Observations	149,286	149,286

*p<0.1; **p<0.05; ***p<0.01

Note: Robust clustered standard errors are reported in parentheses (clusters are the FAP code × Employment agency (ALE) level). These regressions are performed over the whole sample of job seekers (without missing values). The interaction term indicates whether or not there are significantly different treatment effects for individuals with an above median probability/risk of long-term unemployment (LTU), as predicted by an OLS. The significance test for the interaction term with the treatment variable (in bold) indicates whether or not there is treatment effect heterogeneity.

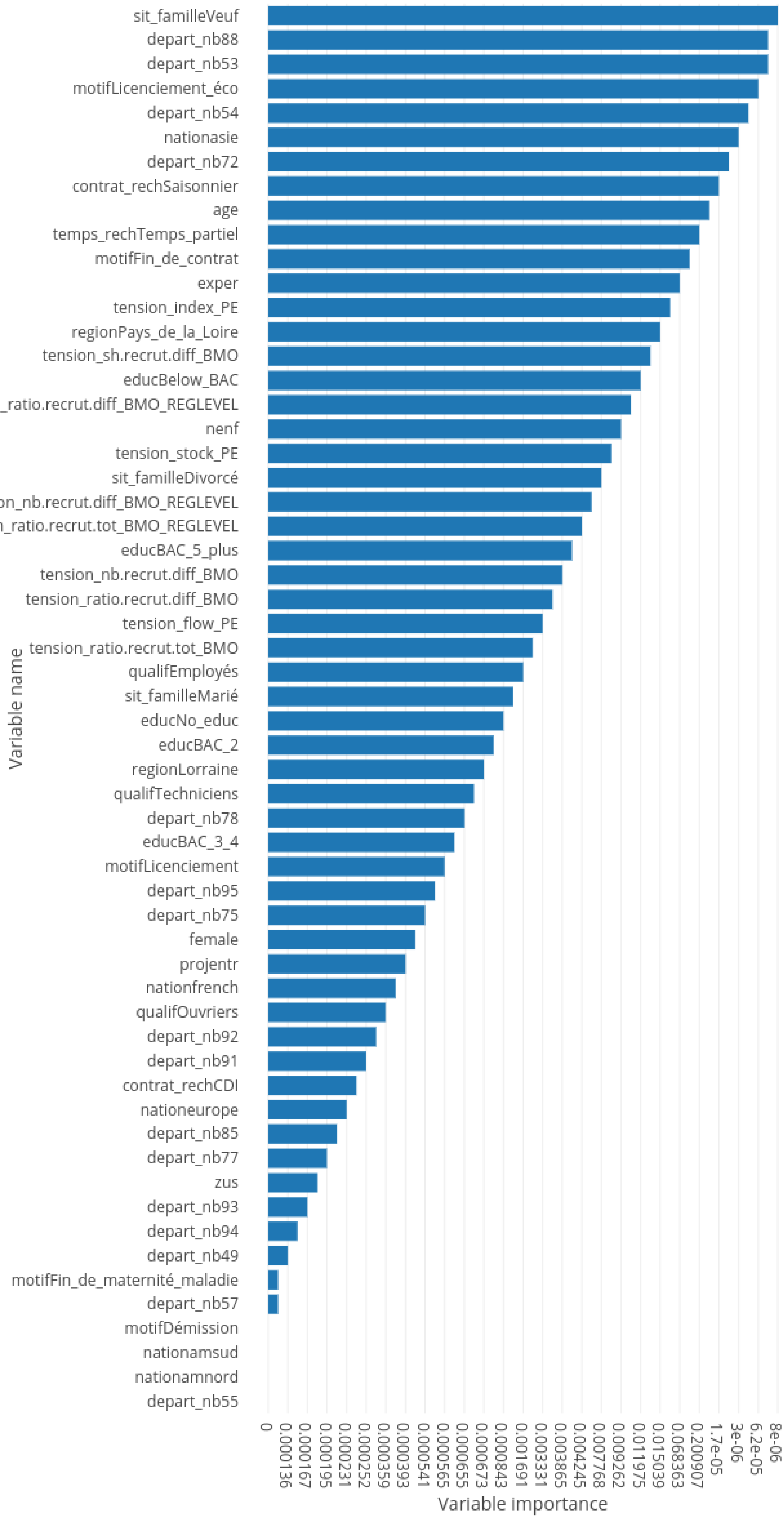


Figure A.2: Variable importance measure for all covariates
— regression forest predicting proba. of LTU

Notes: This figure reports the variable importance measure for all covariates that were candidates for splits in the algorithm that built the regression forest (from `grf` R package "honest" `regression_forest` function) used to predict probabilities of long-term unemployment. This measure is computed using the function `variable_importance` from the `grf` R package.

Table A.6: OLS testing for significant treatment effect heterogeneity across leafs of the tree built on split A and estimated on split B of the data

	Dependent variable:
	Job found after 6 months
leaf-0.0282:group_test	−0.028*
	(0.015)
leaf-0.0256:group_test	−0.026**
	(0.011)
leaf-0.0242:group_test	−0.024*
	(0.013)
leaf-0.0135:group_test	−0.013
	(0.013)
leaf-0.0102:group_test	−0.010
	(0.014)
leaf-0.0048:group_test	−0.005
	(0.012)
leaf-0.0021:group_test	−0.002
	(0.014)
leaf0.0027:group_test	0.003
	(0.015)
leaf0.0036:group_test	0.004
	(0.013)
leaf0.0053:group_test	0.005
	(0.012)
leaf0.0095:group_test	0.009
	(0.014)
leaf0.0099:group_test	0.010
	(0.011)
leaf0.0111:group_test	0.011
	(0.012)
leaf0.0128:group_test	0.013
	(0.008)
leaf0.0129:group_test	0.013
	(0.015)
leaf0.0185:group_test	0.018
	(0.013)
leaf0.0259:group_test	0.026*
	(0.015)
Observations	74,642

*p<0.1; **p<0.05; ***p<0.01

Note: This report the results of an OLS regressing the outcome (job found after 6 months) on the leafs (i.e. interactions of covariates) defined by the causal tree built on split A of the data. This regression is run on split B of the data. I only report the coefficients of the interactions between the treatment and the dummies for each leaf. The name of the leafs indicates the estimate of the treatment effect within the leaf.

Table A.7: OLS testing for significant treatment effect heterogeneity across leafs of the tree built on split B and estimated on split A of the data

	Dependent variable:
	Job found after 6 months
leaf-0.0152:group_test	−0.015
	(0.013)
leaf-0.0137:group_test	−0.014
	(0.013)
leaf-0.0055:group_test	−0.005
	(0.015)
leaf-0.0047:group_test	−0.005
	(0.013)
leaf-0.0044:group_test	−0.004
	(0.007)
leaf-0.0037:group_test	−0.004
	(0.013)
leaf-7e-04:group_test	−0.001
	(0.012)
leaf0.0025:group_test	0.002
	(0.010)
leaf0.0033:group_test	0.003
	(0.009)
leaf0.0195:group_test	0.020
	(0.013)
leaf0.0208:group_test	0.021**
	(0.010)
leaf0.0272:group_test	0.027***
	(0.009)
Observations	74,642

*p<0.1; **p<0.05; ***p<0.01

Note: This report the results of an OLS regressing the outcome (job found after 6 months) on the leafs (i.e. interactions of covariates) defined by the causal tree built on split B of the data. This regression is run on split A of the data. I only report the coefficients of the interactions between the treatment and the dummies for each leaf. The name of the leafs indicates the estimate of the treatment effect within the leaf.

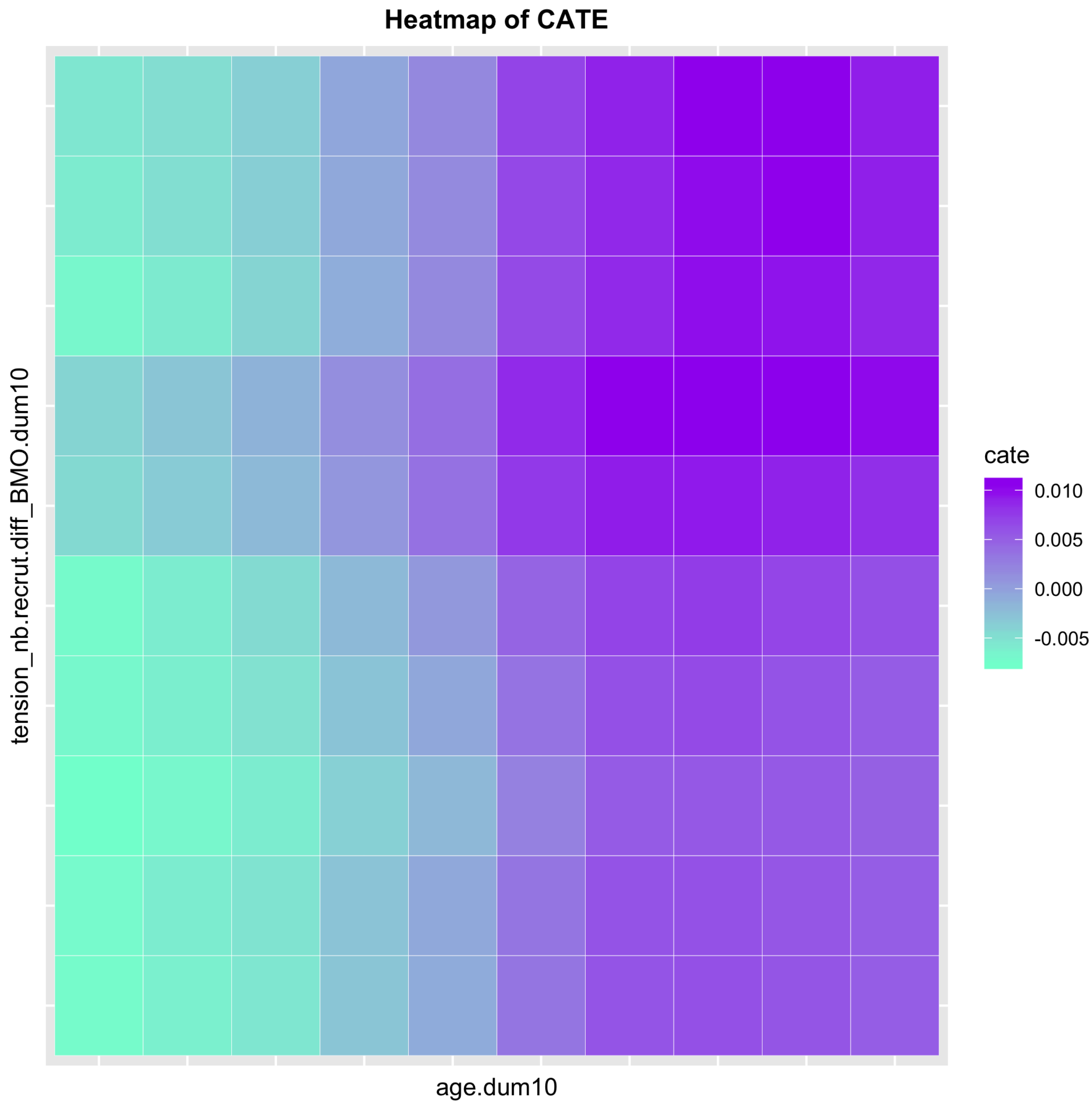


Figure A.3: Heatmap 1 of CATE function

Notes: In this figure, I present a heatmap of the estimated CATE function (see section 6.3.2 for comments). This heatmap plots the number of difficult hirings (BMO) scheduled on the market (deciles) against age (deciles). Each cell color indicates the average value of the CATE function at a given intersection of the two variables plotted.

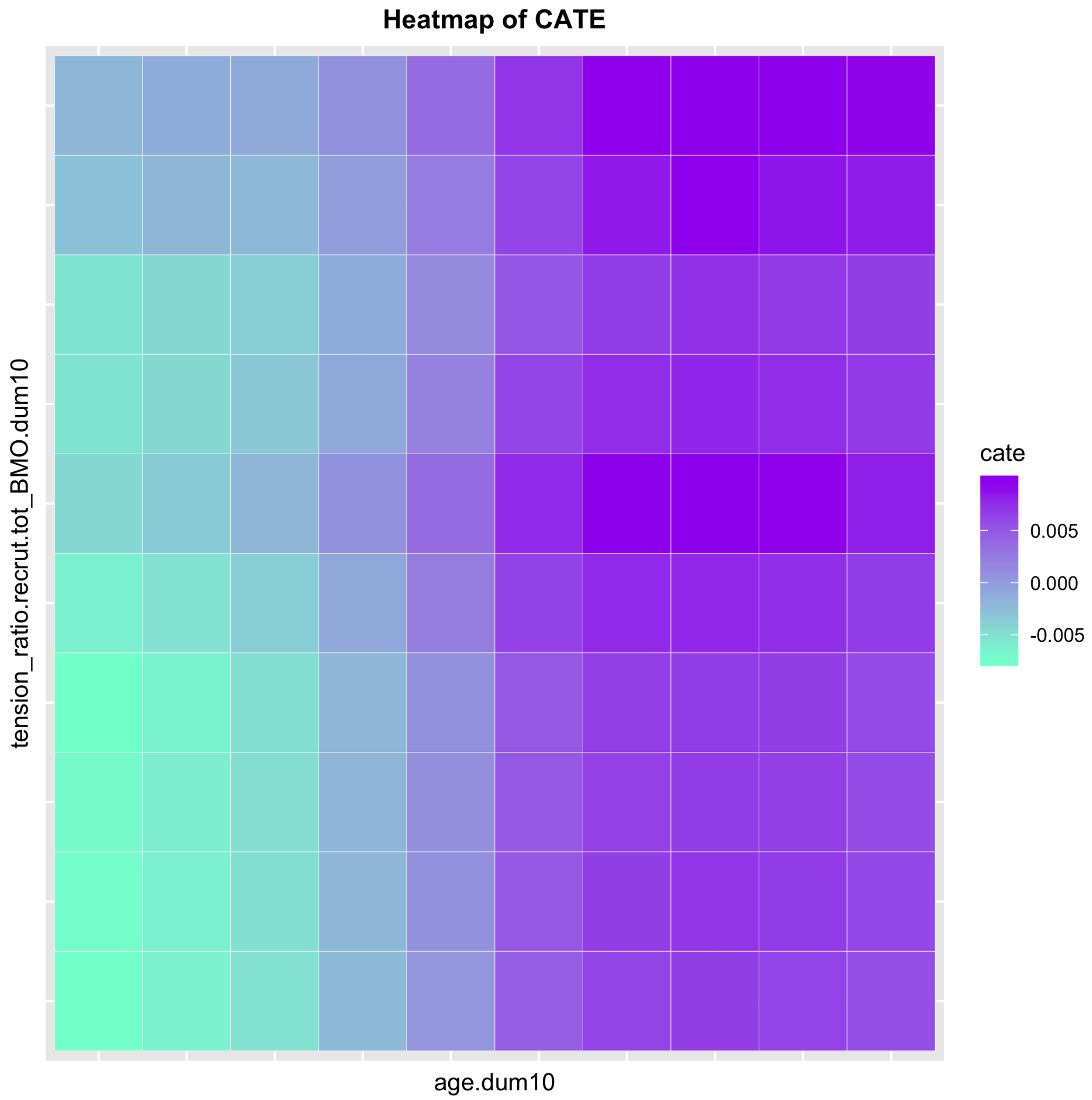


Figure A.4: Heatmap 2 of CATE function

Notes: In this figure, I present a heatmap of the estimated CATE function (see section 6.3.2 for comments). This heatmap plots the ratio of the number of hirings (BMO) scheduled on the market over the (estimated) number of job seekers on this market (deciles) against age (deciles). Each cell color indicates the average value of the CATE function at a given intersection of the two variables plotted.

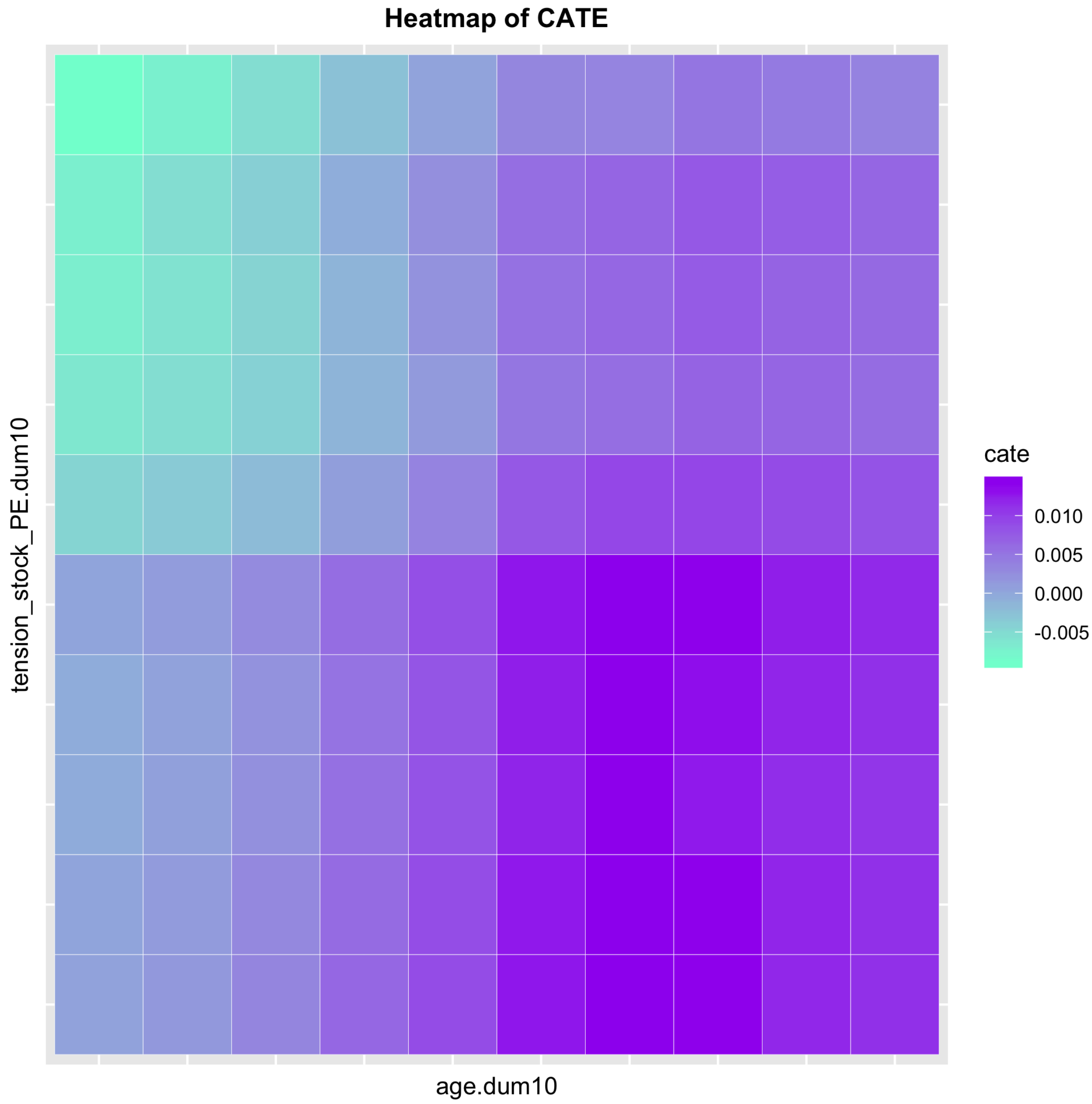


Figure A.5: Heatmap 3 of CATE function

Notes: In this figure, I present a heatmap of the estimated CATE function (see section 6.3.2 for comments). This heatmap plots the ratio of the stocks of vacancies posted at Pôle emploi on the market over the number of job seekers registered at Pôle emploi on this market (deciles) against age (deciles). Each cell color indicates the average value of the CATE function at a given intersection of the two variables plotted.

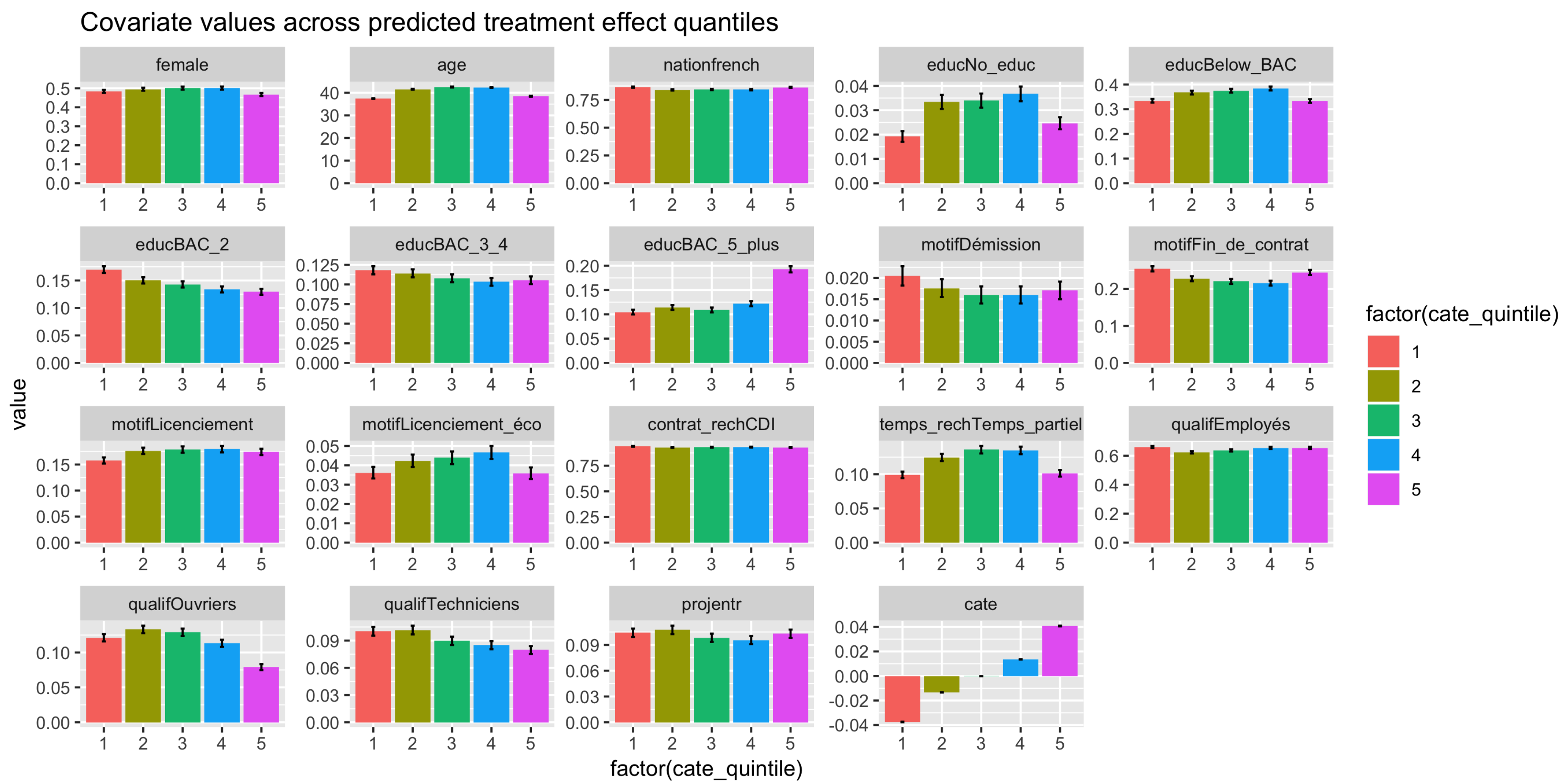


Figure A.6: Avr. values of various covariates along quintiles of estimated CATE function

Notes: In this figure, I present the average values of various covariates along the quintiles of the estimated CATE function (see section 6.3.2 for comments). Quintile 1 and 5 are the quintiles containing respectively the lowest and the highest values of the estimated CATE function.

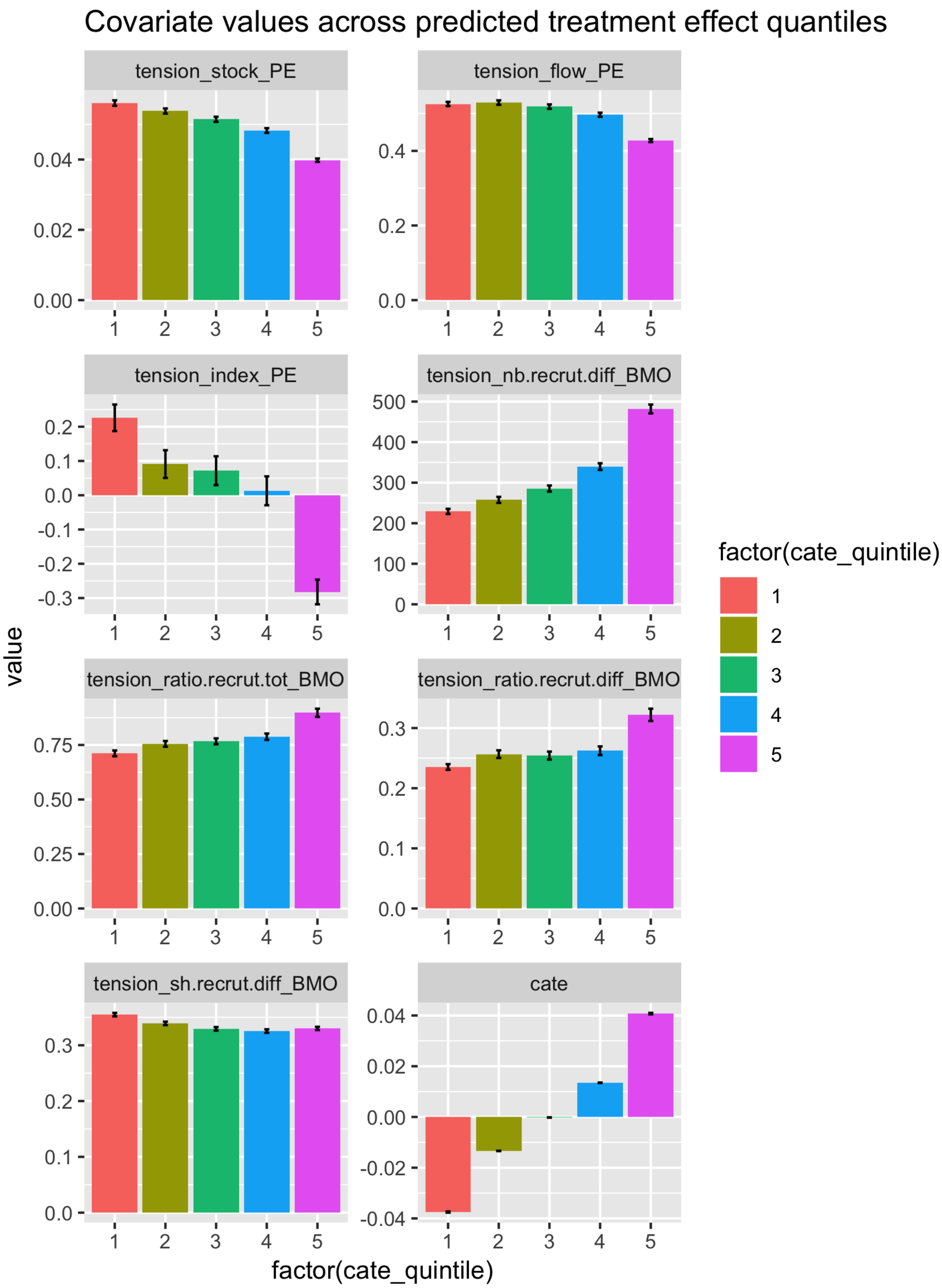


Figure A.7: Avr. values of all tightness indicators along quintiles of estimated CATE function

Notes: In this figure, I present the average values of all available tightness indicators along the quintiles of the estimated CATE function (see section 6.3.2 for comments). Quintile 1 and 5 are the quintiles containing respectively the lowest and the highest values of the estimated CATE function.

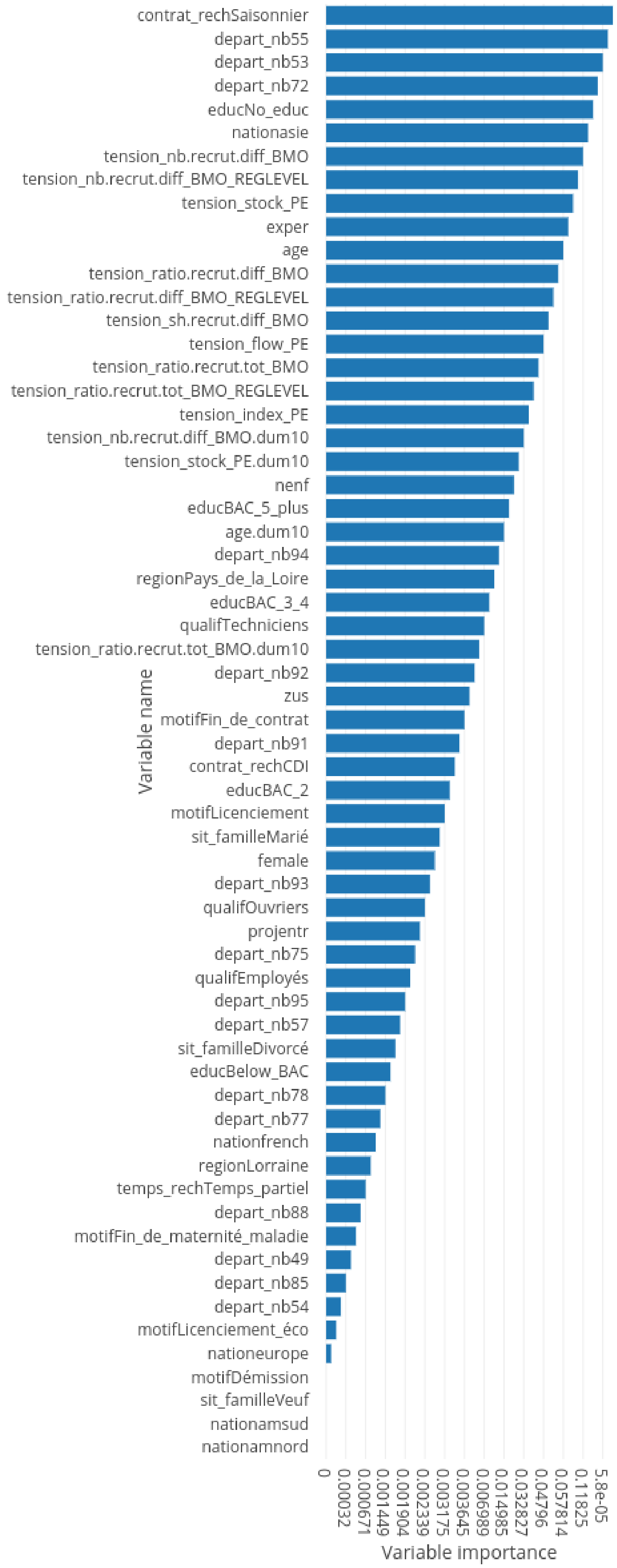


Figure A.8: Variable importance measure for all covariates — causal forest estimating CATE function

Notes: This figure reports the variable importance measure for all covariates that were candidates for splits in the algorithm that built the causal forest (from `grf` R package "honest" `causal_forest` function) used to estimate the CATE function studied in section 6.3.2. This measure is computed using the function `variable_importance` from the `grf` R package.

B Proofs

Proof 1. Let us recall that we have:

$$\frac{\partial^2 p(s,\theta_j)}{\partial T \partial \theta} = \frac{\partial \Delta s}{\partial \theta_j} \cdot \mu(\theta_j) + \Delta s \cdot \mu_{\theta_j}(\theta_j)$$

We want to study the sign of RHS term. If $\frac{\partial \Delta s}{\partial \theta_j} > 0$ then the RHS term is necessarily positive. Considering the case in which $\frac{\partial \Delta s}{\partial \theta_j} < 0$, the RHS term can be positive. The condition for this is:

$$\begin{aligned} \frac{\partial^2 p(s,\theta_j)}{\partial T \partial \theta} &> 0 \\ \frac{\partial \Delta s}{\partial \theta_j} \cdot \mu(\theta_j) + \Delta s \cdot \mu_{\theta_j}(\theta_j) &> 0 \\ \left|\varepsilon_{\theta_j}^{\Delta s}\right| &> \varepsilon_{\theta_j}^{\mu(\theta_j)} \end{aligned}$$

with $\left|\varepsilon_{\theta_j}^{\Delta s}\right|$ the elasticity of the threat effect on s w.r.t. θ_j (which is negative), and $\varepsilon_{\theta_j}^{\mu(\theta_j)}$ the elasticity of the matching rate of applications with vacancies w.r.t. θ_j. □

Proof 2. Let us recall that: $\hat{\mu}(\theta) = \mu(\theta)[1+\varepsilon(\theta)]$. We thus have:

$$\begin{aligned} & \hat{\mu}_\theta(\theta) = \mu_\theta(\theta)\varepsilon(\theta) + \mu(\theta)\varepsilon_\theta(\theta) \\ \Leftrightarrow \quad & \varepsilon_{\theta_j}(\theta_j) > -\frac{\mu_{\theta_j}(\theta_j)\varepsilon(\theta_j)}{\mu(\theta_j)} \end{aligned}$$

□

Proof 3. The first order derivative of this function w.r.t. θ_j is:

$$\frac{\partial s_j^*}{\partial \theta_j} = \left[\varepsilon(\theta_j) + \theta_j \varepsilon_{\theta_j}(\theta_j)\right] \cdot \underbrace{\frac{\gamma c}{zh\nu(1-\gamma)} \cdot \frac{1}{\nu-1} \cdot \left(\frac{\theta_j \varepsilon(\theta_j)\gamma c}{zh\nu(1-\gamma)}\right)^{\frac{1}{\nu-1}-1}}_{>0}$$

Therefore we have:

$$\begin{aligned} \frac{\partial s_j^*}{\partial \theta_j} > 0 \quad & \Leftrightarrow \quad \varepsilon(\theta_j) + \theta_j \varepsilon_{\theta_j}(\theta_j) > 0 \\ & \Leftrightarrow \quad \varepsilon_{\theta_j}(\theta_j) > -\frac{\varepsilon(\theta_j)}{\theta_j} \end{aligned}$$

Let us recall that we assumed:

$$\forall \theta_j > 0, \quad \varepsilon_{\theta_j}(\theta_j) > -\frac{\mu_{\theta_j}(\theta_j)\varepsilon(\theta_j)}{\mu(\theta_j)}$$

And we have:

$$\begin{aligned} -\frac{\varepsilon(\theta_j)}{\theta_j} &< -\frac{\mu_{\theta_j}(\theta_j)\varepsilon(\theta_j)}{\mu(\theta_j)} \\ \Leftrightarrow \quad \frac{1}{\theta_j} &> \frac{\mu_{\theta_j}(\theta_j)}{\mu(\theta_j)} \\ \Leftrightarrow \quad \mu(\theta_j) &> \theta_j \mu_{\theta_j}(\theta_j) \end{aligned}$$

This condition is always verified since we assumed that $\mu(\theta_j)$ is positive, increasing and concave for all positive values of θ_j. □

Proof 4. Let us recall that we have: $s_j^* = \left(\frac{\theta_j \varepsilon(\theta_j) c\gamma}{\nu hz(1-\gamma)}\right)^{\frac{1}{\nu-1}}$
The first order derivative of this function w.r.t. ε is:

$$\frac{\partial s_j^*}{\partial \varepsilon} = \frac{1}{\nu - 1} \cdot \varepsilon(\theta_j)^{\frac{1}{\nu-1}-1} \cdot \left(\frac{\theta_j \gamma c}{zh\nu(1-\gamma)}\right)^{\frac{1}{\nu-1}} > 0$$

□

Proof 5. We have:

$$\underbrace{s_j^*(T=1,\ .\ ) - s_j^*(T=0,\ .\ )}_{(*)} = \left(\frac{\theta_j c\gamma}{\nu hz(1-\gamma)}\right)^{\frac{1}{\nu-1}} \cdot \left(\varepsilon(\theta_j,1)^{\frac{1}{\nu-1}} - \varepsilon(\theta_j,0)^{\frac{1}{\nu-1}}\right)$$

The derivative of $(*)$ w.r.t. θ_j is:

$$\begin{aligned} \frac{\partial s_j^*(T=1,\ .\ ) - s_j^*(T=0,\ .\ )}{\partial \theta_j} &= \underbrace{\left[\frac{1}{\nu-1} \cdot \theta_j^{\frac{1}{\nu-1}-1} \cdot \left(\frac{c\gamma}{zh\nu(1-\gamma)}\right)^{\frac{1}{\nu-1}}\right]}_{(1)} \cdot \underbrace{\left(\varepsilon(\theta_j,1)^{\frac{1}{\nu-1}} - \varepsilon(\theta_j,0)^{\frac{1}{\nu-1}}\right)}_{(2)} \\ &+ \underbrace{\left(\frac{\theta_j c\gamma}{zh\nu(1-\gamma)}\right)^{\frac{1}{\nu-1}} \cdot \frac{1}{\nu-1}}_{(3)} \cdot \underbrace{\left(\varepsilon_{\theta_j}(\theta_j,1)^{\frac{1}{\nu-1}-1} - \varepsilon_{\theta_j}(\theta_j,0)^{\frac{1}{\nu-1}-1}\right)}_{(4)} \end{aligned}$$

The terms (1) and (3) are clearly positive (given that we assumed $\nu > 1$.
Since we assumed that: $\forall \theta_j > 0, \varepsilon(\theta_j,1) < \varepsilon(\theta_j,0)$
and since the function $x \mapsto x^{\frac{1}{\nu-1}}$ is strictly increasing for all positive values of x (given $\nu > 1$), we have that $(2) < 0$.
Lastly, since we assumed that: $\forall \theta_j > 0, \varepsilon_{\theta_j}(\theta_j,1) > \varepsilon_{\theta_j}(\theta_j,0)$
and since the function $x \mapsto x^{\frac{1}{\nu-1}-1}$ is strictly increasing for all positive values of x (given $\nu > 1$), we have that $(3) < 0$.
As a consequence, we have: $\frac{\partial s_j^*(T=1,\ .\ ) - s_j^*(T=0,\ .\ )}{\partial \theta_j} < 0$. □

Proof 6. We have:

$$\frac{\partial p(1,\theta_j)-p(0,\theta_j)}{\partial\theta_j}=\frac{\partial s_j^*(\tau=1,\ .\ )\cdot m_0(1)-s_j^*(\tau=0,\ .\ )\cdot m_0(0)}{\partial\theta_j}\cdot\underbrace{m(1,\theta_j)}_{>0}$$
$$+\underbrace{\left(s_j^*(\tau=1,\ .\ )\cdot m_0(1)-s_j^*(\tau=0,\ .\ )\cdot m_0(0)\right)}_{>0}\cdot\underbrace{m_{\theta_j}(1,\theta_j)}_{>0}$$

Recall that:

$$s_j^*(\tau,\ .\ )=\left(\frac{\theta_j m_0(\tau)c\gamma}{\nu hz(1-\gamma)}\right)^{\frac{1}{\nu-1}}$$

Hence:

$$\frac{\partial s_j^*(\tau=1,\ .\ )\cdot m_0(1)-s_j^*(\tau=0,\ .\ )\cdot m_0(0)}{\partial\theta_j}=m_0(1)\cdot\frac{\partial s_j^*(\tau=1,\ .\ )}{\partial\theta_j}-m_0(0)\cdot\frac{\partial s_j^*(\tau=0,\ .\ )}{\partial\theta_j}$$
$$=\left(m_0(1)^{\frac{1}{\nu}}-m_0(0)^{\frac{1}{\nu}}\right)\cdot\frac{1}{\nu-1}\cdot\left(\frac{\theta_j c\gamma}{zh\nu(1-\gamma)}\right)^{\frac{1}{\nu-1}-1}$$

which is positive since we have $m_0(1)>m_0(0)$ by assumption.
As a consequence, we have that $\frac{\partial p(1,\theta_j)-p(0,\theta_j)}{\partial\theta_j}>0$.

□