An Adaptive Experiment to Boost Online Skill Signaling and Visibility (draft) [‖]

Morgane Hoffmann[*,†], Bertille Picard [‡], Charly Marie [§,†], Guillaume Bied[*,¶]

November 1, 2023

Preliminary Draft. Please do not circulate or cite without permission of the authors. All views expressed in the document are the authors' and not necessarily Pôle emploi's.

Abstract

Digital matching platforms promise to reduce frictions on the labor market by providing low-cost information on available positions and candidates. As such, they may form a welcome addition to the toolbox available to Public Employment Services to bridge labor supply and demand. However, there are certain challenges associated with their adoption. For instance, vulnerable populations may face difficulties in utilizing digital tools effectively. In this study, we evaluate the impact of a communication campaign by email designed to encourage the use of an online matching platform maintained by the French Public Employment Service, Pôle emploi. We design several email templates that combine information, support or motivational content to encourage jobseekers to engage with their profiles on the platform. To assess email effectiveness, we implement an adaptive experiment (contextual bandit) aiming to leverage past jobseekers' take-up responses and characteristics to determine future email allocation, gradually reducing the allocation of less promising templates. Additionally, we built an optimal personalization allocation strategy based on collected data and test its effectiveness. Emails has a positive impact on the usage of the platform, as measured by a wide range of outcomes. However, attempts at learning a personalized emailing strategy do not manage to significantly improve on a random allocation of email templates.

Keywords: Online Job Platforms, Contextual Bandits, Digital Tool Adoption

[*]Centre de Recherche en Économie et de Statistiques (CREST), CNRS, École polytechnique, GENES, ENSAE Paris, Institut Polytechnique de Paris, 91120 Palaiseau, France

[†]Pôle emploi

[‡]Aix-Marseille University, CNRS, AMSE, 13000 Marseille, France, Corresponding author bertille.picard@univ-amu.fr

[§]University of Lorraine (PErSEUs), Île du Saulcy, 57045 Metz, France

[¶]Université Paris-Saclay, CNRS, Laboratoire interdisciplinaire des sciences du numérique, 91405, Orsay, France

[‖]We would like to thank David Bourguignon, Bruno Crépon, Laura Litvine, Molly Offer-Westort, Vianney Perchet, Elia Perennes, Roland Rathelot and Arne Uhlendorff for comments and suggestions. This work would not have been possible without the support of Pôle emploi; we particularly acknowledge help and comments from Catherine Beauvois, Ludivine Degand, Cyril Nouveau, Jean-Pierre Tabeur, Matthieu Teachout, Chantal Vessereau and essential technical support from Axel Gaugler.
This project was pre-registered on the AEA's RCT Trial Registry and can be accessed through https://www.socialscienceregistry.org/trials/10085 (Bied et al. 2023). It was approved by the Paris School of Economics' Institutional Review Board.

Contents

1 Introduction

Over the past decade, the internet revolution has transformed the recruitment landscape by providing individuals with a variety of new tools (Kuhn 2014). The emergence of online job boards such as Indeed, LinkedIn, Glassdoor, and Monster in the private sector has created centralized platforms where jobseekers and recruiters can connect with ease. These innovations hold promise in reducing information asymmetries between labor supply and demand (Wheeler et al. 2022, Jones and Sen 2022).

Online platforms offer a range of features, including the ability for jobseekers to create online profiles to signal their skills and qualifications. These profiles can be easily accessed by recruiters, enhancing visibility and increasing the chances of being discovered by potential employers. On the one hand, some experimental studies have demonstrated the potentialities of credible information on jobseekers' skills or past performance in improving labor market outcomes (Carranza et al. 2021; Pallais 2014; Bassi and Nansamba 2018-01; Abel et al. 2020; Abebe et al. 2020). However, it is important to note that low-cost access to information on these platforms does not always guarantee better results at the micro-economic level (Dhia et al. 2022). Part of the explanation may lie in the behaviors of online actors, as highlighted by the study by Marinescu and Wolthoff (2020) on the recruiters' side. However, the behavior of individuals on these platforms remains little studied. This paper aims to partially fill a gap by examining how to encourage job seekers to report their skills online, and to bring in new users. More specifically, we focus on removing barriers that might prevent them from using online tools.

To leverage the potential advantages of online resumes, the French Public Employment Service (PES), *Pôle emploi*, offers an online service called the *Profil de Compétences* (referred to as skills profile). It enables jobseekers to post their resumes, skills as well as past professional achievements online and to be viewed by recruiters. However, the adoption of this platform among jobseekers remains low, with only approximately 20% of jobseekers publishing their profiles. This issue of low adoption of public services is a well-documented challenge in the literature (Baicker et al. 2012; Bhargava and Manoli 2015; Janssens and Van Mechelen 2022). It is crucial to understand the causes of this non-take-up and develop effective incentives to encourage jobseekers to engage with digital tools that can be beneficial to them. In this study we partnership with the PES and design different emailing incentives aimed at increasing jobseekers' engagement with the online skills profile. Our interventions are informed by previously identified factors, such as Bhargava and Manoli's (2015) "psychological frictions" that can hinder individuals from fully utilizing public assistance. Based on these insights, we aim to specifically address and mitigate the impact of similar factors to enhance individuals' willingness to engage with the service. Previous experiments have shown a positive response of jobseekers to emails sent by the French public employment agency (Dhia et al. 2022, and Dhia and Mbih 2020 with lower but significant response). We question whether individuals may face barriers such as a lack of knowledge and understanding of the tool, suitable opportunities, limited capacity to engage with the platform, or simply a failure to perceive the advantages of using it (Michie et al. 2011). Because these barriers can vary among jobseekers, we have designed 16 different emails tailored to tackle these different causes of non take-up while taking into account that different jobseekers may respond differently to these incentives.

To test the effectiveness of these various incentives on jobseekers' engagement and then explore heterogeneity of responses by subgroups, one could randomly send incentives to jobseekers. However, conducting a standard randomized control trial for our large set of potential email types can be costly, as a substantial number of observations is needed to accurately estimate their effects. This cost increases even further when considering the effects of treatments within subgroups. To address this concern, we propose using an adaptive experiment, which offers an efficient approach to test the different incentives by collecting observations and focusing on the most promising treatment arms. In this design, the value of information obtained from the most promising arms is prioritized. This means that researchers choose to gain more knowledge about the treatment arms that show the most potential, while allocating less statistical power to explore less promising arms. This design can also be considered a more ethical one from the perspective of experimental subjects. Individuals have a higher probability of being assigned to a favorable treatment, compared to a randomized control trial where the treatment assignment is purely random. We contribute to the growing literature on applications of contextual bandit methods in the context of randomized control trials in economics (Caria et al. 2020), political science (Offer-Westort et al. 2021) or clinical trials (Villar et al. 2015).

Our experiment shows a positive and significant effect on the probability of jobseekers visiting the skills profile. This effect was observed across various measures of take-up, including repeated visits to the platform and modifications to the profile. Most affected jobseekers are mainly individuals who were closer to employment and to the institution. However, we observe no increase of the treatment effect during the experiment, meaning that the use of a contextual bandit strategy failed to improve on the baseline of random email assignment. In a last phase of the experiment, we compare a personalized allocation strategy based on a policy tree against a random allocation and find no significant difference.

The paper is organized as follows. In Section 2, we provide an overview of the experiment's context, including a description of the target population and the interventions implemented. Section 3 details the experimental design employed to assess the effects of the interventions. In Section 4, we present the findings of the global mailing campaign, highlighting its impact on jobseekers' engagement. Section 5 examines the relative effectiveness of different incentives in driving engagement with the skills profile. In Section 6, we evaluate the effectiveness of a learned personalized allocation policy compared to a random allocation, investigating whether personalization improves outcomes. We conclude the paper in Section 7 by discussing the results and the implementation of the design.

2 Context, Diagnosis, and Interventions

2.1 Context

The French Public Employment Service, Pôle emploi, is an intermediary between recruiters and jobseekers. It is responsible for a number of tasks for jobseekers, such as distributing unemployment benefits, monitoring them and helping them in their job search. The PES introduced the skills profile in June 2018. This service allows jobseekers to create and publish a detailed online profile on its platform.[1] Jobseekers are invited to complete their profile in their personal space, as soon as they register with the institution. Users can also choose whether to make their profile visible to recruiters or only to caseworkers.

The profile was conceived to promote a "skills-based approach", encouraging job seekers to document and put forward their "hard" and "soft skills" which may be transferable across occupations. Users can enter the following information on the platform : (1) career path, including professional experience, training and areas of interest, (2) skills, such as spoken languages, driving licenses and hard and soft skills, (3) resume and achievements, including a PDF resume, diplomas and certifications, and (4) the sought occupation and/or professional projects. This information is summarized in a "card" that is displayed to employers on the platform. Employers can click on the card to view the full profile and contact the jobseeker.

2.2 Identifying Barriers to Profile Take-up using Behavioral Science

We rely on a behavioral science lens to better understand people's low take-up of the profile and ways to encourage its use (Hallsworth 2023) Works that have used behavioral science to address non-take-up by sending communications point to mixed results depending on the public provision, time frame, targeted and reached population, or addressed "psychological frictions." Building on this literature, previous experimental studies and internal workshops, we propose three main types of barriers to use of the profile: lack of information and understanding of the tool and ways to fill it out, complexity of the tool, low motivation to do so because the person might not perceive the advantages of using it. Following Michie et al. (2011), we organize this ideas as *capacity* factors, *opportunity* factors and *motivational*.

Capacity factors. We hypothesize that the first of these barriers stems from a lack of capacity, defined as one's psychological and physical capacity to perform a given behavior, due in part to (1) a lack of knowledge about the mere existence of the tool, and (2) a lack of skills to fill it out. Internal workshops revealed that some users may not understand the concept of skills, may not have the ability to effectively utilize a skills-based approach, by creating and filling in a skills profile. This is consistent with previous research showing that non-take-up could be due to lack of information (Bhargava and Manoli 2015, Finkelstein and Notowidigdo 2019). These issues could lead to users not engaging with the tool due to a lack of understanding and a low feeling of self-efficacy and capacity (Bandura 1997)[2]. In addition, the tool is not anchored to an existing representation (Moscovici 2008, Wagner et al. 1999). Not being spontaneously associated with a more familiar object, such as a resume, might hinder its uptake (Schanzenbach 2009). Finally, the lack of systematic promotion of the skills profile during the initial interview could both affect the mere knowledge of its existence as well as judgments about it and its future use (Dechêne et al. 2010, Tversky and Kahneman 1973-09).

Opportunity factors. Opportunity factors lie outside the individual and make a given behavior simpler or more complex. Here, the tool's complexity may exceed users' cognitive abilities (Dia-

[1] https://www.pole-emploi.fr/accueil/

[2] Self-efficacy is a malleable and context-dependent belief that a person holds about one's (1) ability to perform a specific behavior and (2) one's control over one's environment.

mond 2013), which may in turn undermine its take-up (Bhargava and Manoli 2015, Finkelstein and Notowidigdo 2019, Linos et al. 2022, Schanzenbach 2009). Moreover, the large number of available choices and options could overwhelm people, who may be in a situation of *choice overload* and thus not use the profile (Chernev et al. 2015, Scheibehenne et al. 2010-10), especially among low-income users (Baicker et al. 2012, Bhargava et al. 2017, Mani et al. 2013). These hypotheses are supported by internal studies.

Motivational factors. Motivational factors are processes that energize and direct a given behavior. The tool could either diminish users' motivation because of its complexity and the costs associated with its use (Bhargava and Manoli 2015, Linos et al. 2022, Schanzenbach 2009), or not foster user's motivation because of a lack of information and a low feeling of self-efficacy (Bhargava and Manoli 2015, Finkelstein and Notowidigdo 2019).

First, the time and effort required to complete the profile is substantial in the short run, but the benefits of a well-completed profile may only materialize in the medium or long run. This might lead people to postpone its completion, because of a *time inconsistent* or *present bias* preference (Baicker et al. 2012, Laibson 1997). Moreover, the very short-term rewards of filling it out might be perceived as low, if not nil, which could again lead to postpone its completion (Woolley and Fishbach 2017, Woolley and Fishbach 2018).

Then, uncertainty about the consequences of a well-filled profile might also contribute to its non-take-up (Dynarski et al. 2021-06).

Moreover, and more classically, the cost-benefit ratio of the tool appears to be negative. This is reflected in the fact that one out of three jobseekers does not consider that the profile has a positive impact in terms of better identifying their skills, on the targeting of job offers and on the facilitation of exchanges with their caseworker.

Finally, difficulties in using the tool may reduce the user's self-efficacy, which could then weaken his or her motivation (Bandura 1997, Deci and Ryan 2000, Ryan and Deci 2000).

2.3 Interventions Design and Rationale

Following our diagnosis and building on previous work, we designed emails to encourage jobseekers to engage with their profile (i.e to fill it and publish it on the platform to become visible to recruiters). Treatments address our 3 main barriers of a lack of capacity, opportunity and motivation by either providing additional information or by attempting to reduce the perceived cost of using the tool. Further details about mails variations are presented in appendix 1.

Our variations can be divided into 3 types of levers, each of which has an empty control baseline: basic presentation of the tool, motivating the use of the tool, providing help to use the tool. In addition, two sending hours are possible with the email being sent either in the morning at 9am or the afternoon at 3pm. Figures 1 and 2 provide an overview of the email structure and of possible combinations, to better understand the email structure. Appendix C provides a detailed description of the contents of the emails used in the experiment and Appendix C.4 presents an example of one of the emails that was sent to jobseekers.

General email structure. All email share the following fixed general structure:

- **To get the users' attention, we use their name at the beginning of the email**

  Because people's attention is limited (Bordalo et al. 2022, Chun et al. 2011), we strive to capture the person's attention by using their name at the very beginning of the email, an effect known as the "cocktail party effect" (highly relevant stimuli, like hearing one's name at a noisy party, can capture people's attention, Conway et al. 2001).

- **A brief presentation of the profile**

  This enables us to address the lack of information about its existence and purpose as well as the lack of information about what skills and a skills-based approach are. To ensure that jobseekers understand the main purpose of the tool, we again follow Schanzenbach's work (2009) and anchor the skills profile in a CV representation (Moscovici 2008, Wagner et al. 1999).

- **A reminder to update the profile**

  Some jobseekers may have already published their profile, but may have not updated it. We remind them that it is useful to update their profile, even if it has already been published.

Informational treatments

- **I1 : Information regarding recruiter usage**

  To emphasize the potential benefits of filling out one's profile and to motivate users to do so, we highlight its use by other agents. Indeed, it is important that the platform appear attractive when there are private competitors on the market such as LinkedIn or Monster. Specifically, we provide information about the number of recruiters who use the platform to search for candidates each month ($\approx$ 350,000). In doing so, we highlight its potential benefits to motivate its use in a classic cost-benefit argument (Bhargava and Manoli 2015).

- **I2 : Information about improvements in service quality**

  Caseworkers rely on these profiles to recommend job ads to jobseekers, and automated recommender systems built by the institution also use them. A well-completed skills profile can enable Pôle emploi's caseworkers to better communicate with jobseekers and better assist them in their job or training search. We highlight this information to jobseekers to capitalize on the trust they have towards their caseworker, which has been shown to be a precursor to cooperation (Balliet and Van Lange 2013).

Help provision treatments

- **H1 : Including an autonomous help device**

  To reduce felt completion costs and increase the feelings of autonomy and self-efficacy, we provide a step-by-step guide of the tool directly in the body of the email (Bandura 1997, Deci and Ryan 2000, Ryan and Deci 2000). We do so because previous works have shown that the effect of providing information is greater when combined with assistance (Bettinger et al. 2012, Finkelstein and Notowidigdo 2019).

- **H2 : Including an intensive help**

  Beyond using the tool in autonomy, some jobseekers may find it difficult to complete their profile on their own due to difficulties with the French language, digital tools, a lack of self-efficacy or other reasons. To address these challenges, we offer intensive support in two forms:

  - **H2w : Workshop**

    We offer jobseekers to register for a half-day workshop to help them fill in their profile on the PES website. Jobseekers only need to log in and will then be taken directly to the registration page.

  - **H2c : Caseworker**

    We offer to schedule an appointment with the caseworker if the user needs personalized assistance, again to leverage the trust towards the caseworkers. We do this through a call-to-action button that also requires the user to log in and then be taken directly to the appropriate appointment setting page.

- **S : Sending hours**

  We send emails at two different times: 9am or 3pm. We vary the sending times because we expect that candidates' availability may vary throughout the day depending on their characteristics. For example, having a job or having children may influence the time at which the email is read.

It is important to note that all jobseekers have access to the online tool, and our intervention does not prevent any jobseeker from benefiting from the service. Jobseekers are free to create, modify or publish a profile whether assigned to the control or the treatment group.

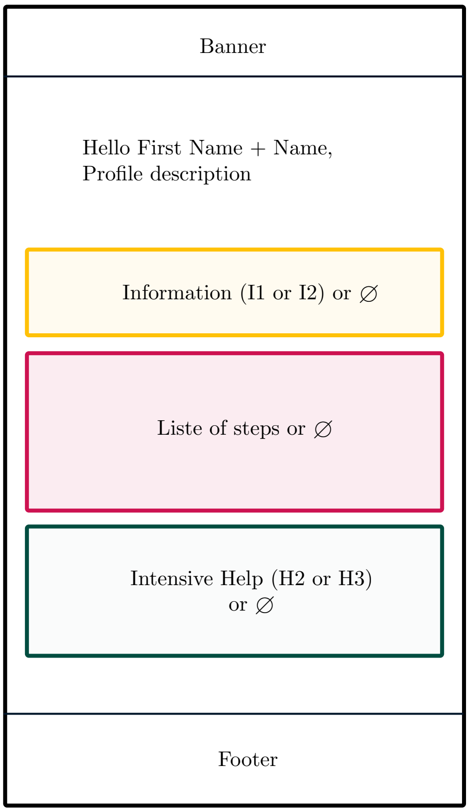


Figure 1: Generic overview of an email

Note: "Information", "Tutorial" or "Intensive Help" may appear or not (see figure 2)

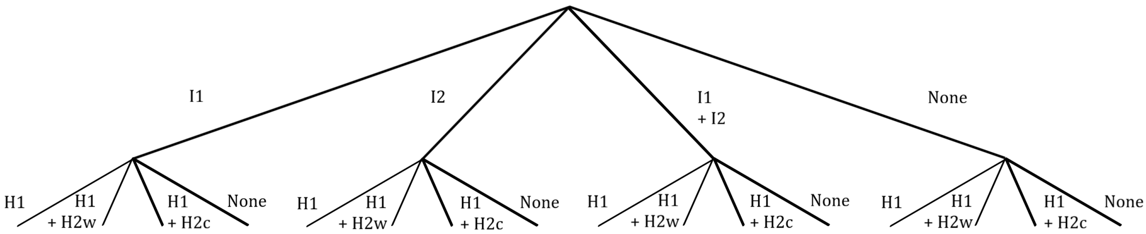


Figure 2: Possible combinations of blocks

3 Experimental design

3.1 Rationale for Adaptive Experimentation and Goals

Our objective is to evaluate the effectiveness of mailing interventions in increasing the uptake of the skills profile among jobseekers. We define uptake as the probability to visit the profile's webpage. Additional information regarding the rationale for choosing this measure of success are given in 3.4.4. We want to know which emails work best for whom in that regard, in order to later on learn an efficient contextual policy. Since we think filling the profile should benefit job seekers and the PES, maximizing the profile uptake during the experiment is also one of our aims.

To achieve these goals, we use a contextual adaptive experiment design. The algorithm's objective is to reconcile two contradictory goals: sending the email that appears most promising to the individual based on collected data while gathering more information about emails whose effectiveness remains uncertain, which is known as the "exploration vs. exploitation" trade-off. This approach allows us to dynamically adjust the treatment allocation based on the accumulated data. It enables us to optimize the impact of the interventions on user engagement with the skills profile during the experiment.

We denote the treatment arms $w \in \mathcal{W}$, with $\mathcal{W} = \{1, ...K\}$ where $K = 32$ in our setting. Individuals are divided into batches that are successively drawn. Let us the set of batches $\mathcal{B} = \{\mathcal{B}_1, ...\mathcal{B}_B\}$. Thus, $\mathcal{B}_1$ contains all the indices of the individuals drawn during the first round. $Y_i(w)$ is the potential outcome, that is, based on Rubin's framework (1974), the outcome individual i sent to treatment arm w would receive. These potential outcomes follow unknown distributions that may differ from one arm to another or from one individual to another due to heterogeneity. We denote $\mu_i^w = \mathbb{E}\big[Y_i(w)\big]$ the expectation of the potential outcome[3]. This allows us to define the expected regret[4] after $T \in [\![1, B]\!]$ batches as:

$$R_T = \sum_{b=1}^{T} \sum_{i \in \mathcal{B}_b} \left(\mu_i^* - \mu_i^{w_i}\right)$$

with $\mu_i^* = \sup\{\mu_i^w : 1 \leqslant w \leqslant K\}$ and w_i being the treatment arm that has actually been assigned to i. This is the (expected) cost paid after allocating sub-optimally treatment arms. The typical overall goal of bandits is to minimize this quantity. In the experimental context, we know neither the full distribution of the potential outcomes nor their expectations for each individual in advance. Moreover, μ_i^* cannot be estimated easily since an individual is observed in one treatment arm only.

We can ask whether the allocation strategy is improving over time by calculating the averages of observed rewards, either in each arm, or averaged over all treatments. Let $n_{k,T}$ be the number of times the k arm has been drawn after T batches:

$$n_{w,T} = \sum_{b=1}^{T} \sum_{i \in \mathcal{B}_b} \mathbb{1}\big(\{w_i = w\}\big)$$

[3]Please note that there is no empirical counterpart to this quantity: $Y_i(w)$ is never fully observed on the population of experimental subjects since some individuals have been allocated to treatment arms different from w. In particular, this quantity is distinct to the expected reward given that the treatment arm w has been selected. If the reader is used to the notations in the reinforcement learning literature, for example those used by Sutton and Barto (2018), they could note that μ_i^w here is different from $q_*(w) = \mathbb{E}\big[Y_i^{\text{obs}} \mid W_i = w\big] = \mathbb{E}\big[Y_i(w) \mid W_i = w\big]$ which is the average reward among people receiving the treatment arm w. Due to selection bias induced by the adaptive algorithm, we can expect $q_*(w) \neq \mu_i^w$. While working on non-observables or counterfactuals may be less common in the reinforcement literature, hence the conditioning on the event $W_i = w$, it is very common in the experimental literature and we choose to adopt this convention here. This will facilitate our description of the empirical strategy later.

[4]We base our notations here on Wager's lecture notes on causal inference (2022), with some minor changes to adapt to the contextual case. In this document, Wager rephrased the bandit problem to align more closely with the notations commonly employed in the literature on causal inference, reflecting the potential outcomes framework experimenters are familiar with.

The current average of rewards associated to the treatment arm w after T batches is then:

$$\hat{m}_{w,T} = \frac{1}{n_{w,T}} \sum_{b=1}^{T} \sum_{i \in \mathcal{B}_b} y_i^{\text{obs}}$$

where y_i^{obs} is the empirical counterpart of $Y_i^{\text{obs}} = \sum_{w \in \mathcal{W}} Y_i(W_i)\mathbb{1}\{W_i = w\}$, that is the observed value taken by the random variable equal to the potential outcome associated to the treatment arm that is allocated.

In order to discover this optimal allocation strategy we use a contextual bandit algorithm that balances between exploration and exploitation. Following Dimakopoulou et al. (2018), we use a variant of the Thompson sampling procedure (Thompson 1933). This algorithm requires us to estimate the expected rewards for individuals of a given batch, which is described below.

3.2 Learning the Expected Rewards during the Experiment

To better decide how to allocate treatment arms to individuals in a batch, we need to evaluate the potential gains or rewards associated with each of these arms. For each batch, this value can be predicted on the basis of data collected in the past, i.e. observations from previous batches.

Thus, we will assume that differences between individual distributions of potential outcomes are strongly correlated with differences in individual characteristics $x_i \in \mathcal{X}$, where $\mathcal{X}$ denotes the support of the vector of individual characteristics. Therefore:

$$\forall i, \quad \mathbb{E}\big[Y_i(w) \mid X_i = x\big] = \mu_w(x)$$

In other words, to predict potential outcomes, we will compare the individual with individuals who resemble him or her in terms of characteristics X in the available database. The estimation of $\mu_w(x)$ for all $w \in \mathcal{W}$ and $x \in \mathcal{X}$ could therefore be performed using any predictive model[5].

Following Dimakopoulou et al. 2018 estimate non-parametrically $\mu_w(x_i)$ using an honest random forest model (Athey et al. 2019), predicting the observed outcomes y_i^{obs} in the past using the information on the individual characteristics x_i, the allocated treatment arms w_i and the detailed components of the email, also referred to as "behavioral and informational levers" in the continuation of this paper. Including dummy variables for these levers allow us to take into account potential correlation between treatment arms sharing some components. This model allows us to obtain estimates of $\hat{\mu}_w(x_i)$ and its variance $\hat{V}(\hat{\mu}_w(x_i))$ based on the past collected data[6]. The choice of this algorithm is based on theoretical arguments and on the context of the experiment. As soon as we collect the first batch, we will have already a very large number of individuals at our disposal to estimate the rewards (12 000 divided in 32 groups). At the same time, administrative data offer the possibility of using numer-

[5] In fact, since the algorithm's goal is mainly to allocate the treatment arms that are *relatively* the best, missing variables that determine all potential outcomes in the same direction and the same proportion could affect the estimation of $Y_i(w)$ without hurting the identification of the best treatment arms. It is the case when the missing variable is not a confounder. In other words, the above assumption can be relaxed a bit, assuming that for any pair (w, w'), $\mu_i^w - \mu_i^{w'} = \mu_w(x) - \mu_{w'}(x)$.

[6] This variance is valid thanks to the honesty property. The honesty property is achieved here through sample splitting as described in Athey et al. (2019). The estimation procedure for the variance is described in section 4 of Athey et al. (2019).

ous individual characteristics. We have no simple hypotheses on the functional relationship between potential outcomes and personal information, hence the choice of a non-parametric approach. This explain why we do not considerate as an alternative model imposing a linear relationship between the context and the reward, $\mu_w(x_i) = x_i^T\theta_w$, and estimating the parameter θ using ordinary least squares. With a generalized random forest, we aim to capture potential complex relationships in "real data." Moreover, we leverage the honesty property of generalized random forests, which ensures that the sample used to construct the trees is independent of the sample used to estimate the improvement in fit resulting from a split. This honesty property helps to mitigate bias and overfitting issues, particularly in the early stages of the learning process.

Using an honest random forest model comes with the drawback of potentially slower identification of the most effective treatment arms, which can be attributed to three main factors. Firstly, unlike models with more specific specifications like linear regression, the random forest lacks structure, which may result in slower decrease of the regret. Secondly, honesty here is achieved through sample splitting. Lastly, the number of variables included in the model can also affect convergence, as a higher number of variables increases complexity and may hinder efficient convergence. This trade-off can be viewed as a design choice that gives more chances to exploration over exploitation.

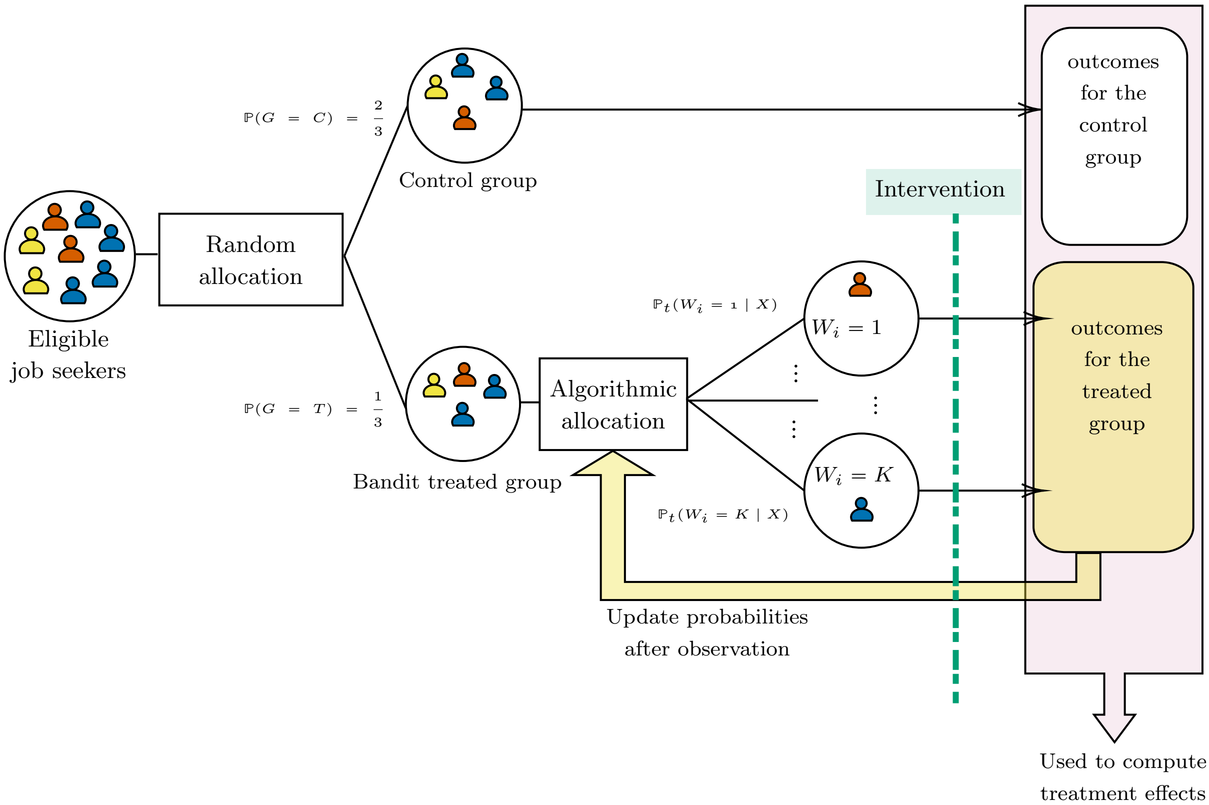
3.3 Overview of Allocation Procedure

In this section, we provide a brief overview of the dynamic procedure followed to allocate jobseekers to different treatments. For a more comprehensive understanding of the allocation process, please refer to Appendix G.1 where detailed information is provided.

Thompson sampling procedure At each round, we want to estimate the conditional expected gain for each arm $w \in W$ and for each individual given its context x_i, $\mu_w(x_i)$. In order to estimate this quantity one could use any supervised machine learning algorithm. We use at each round all the data collected $\{X_i, W_i, Y_i\}$ to obtain an estimate $\hat{\mu}_w(x_i)$ and $V(\hat{\mu}_w(x_i))$ for all arms w. We assume that the expected gain for arm w, $\mu_w(x_i)$ given context x_i follows a normal distribution: $\mathcal{N}\left(\hat{\mu}_w(x_i), \hat{V}(\hat{\mu}_w(x_i))\right)$. Eventually, we draw samples from the distribution for all arms w and choose the arm that has the highest value. The choice of drawing the reward from a normal is in line with the existing literature on contextual Thompson Sampling, both older (Agrawal and Goyal 2013) and recent (Dimakopoulou et al. 2018). While a misspecification of the reward distribution cannot be ruled out, it shall have little effect on the algorithm's behavior. For example, Agrawal and Goyal (2013) emphasize that in the context of linear contextual Thompson Sampling with continuous rewards, the regret bounds will hold irrespective of whether or not the actual reward distribution matches the Gaussian likelihood function used. Insofar as what matters first is the frequency with which the arm gives the highest draw among the draws of each arm, modeling these draws by the wrong distribution will at most have an influence on exploration, although this remains controlled by the estimated variance. However, the use of a normal distribution rather than a beta distribution, which might seem more appropriate here as the reward is a probability, in the non-linear context could be further investigated theoretically.

An important modification made to the allocation procedure is that we define a probability floor of 0.005 to ensure that the probability of being assigned to any treatment is bounded away from zero or one. This minimum probability ensures that we keep exploring all the treatments during the experiment, even if it appeared to perform poorly in the first periods. It is also preferable to keep non-zero sampling probabilities for the inference.

Figure 3: Experimental design



Note: Emails are sent every Tuesday at 9am or 3pm, data is collected the next Monday at 6pm in order to run the algorithm for the next wave

Experimental Procedure The experimental design follows a consistent procedure each week, as illustrated in Figure 3. The steps involved are as follows:

1. Observe and collect responses from individuals in the previous batch.

2. Augment the training dataset by including the bandit treated group responses.

3. Update the predictive model using this augmented dataset.

4. Sample 36,000 jobseekers at random from the eligible population.

5. Divide the sample of 36,000 jobseekers into two groups: the control group, consisting of two-thirds of the sample (24,000 individuals), and the bandit treated group, consisting of one-third of the sample (12,000 individuals).

6. For each individual in the bandit treated group and for each treatment, predict the conditional expected gain using the updated model.

7. Apply the Thompson sampling procedure, derive individual treatment allocations for each individual in the new sample and assign these individuals to treatment.

8. Send emails to the assigned individuals.

This procedure was repeated seven times during the main experiment, conducted from January 10th to February 28th, 2023. As a result, we obtained a treated sample consisting of 84,000 individuals and a control sample consisting of 168,000 individuals[7].

3.4 Data Sources and Sample Characteristics

3.4.1 Data Sources

The primary data source used is the administrative record maintained by the PES. This record contains essential socio-demographic information such as age, gender, education level, and work experience. It also includes details about the jobseekers' unemployment history.

Another important source of information is the log data from the profile webpage. This log captures the various actions taken by jobseekers on the website, particularly on the pages related to the profile service.

By linking the log data with the jobseekers' national identification number, we can merge the log data with individual characteristics.

3.4.2 Inclusion Criteria

Our eligible sample consists of jobseekers who are registered with the French Public Employment service and reside in metropolitan France. Only jobseekers for whom we do not have a registered email address, or who have not consented to receive emails about Pôle emploi's services, are excluded from the experiment. jobseekers who meet these criteria are included in the experiment, regardless of whether they already have an existing profile or not. Our sample is representative of the population of registered jobseekers in France.

3.4.3 Covariates

We consider a comprehensive set of covariates, encompassing socio-demographic variables, targeted job preferences, characterization of experienced unemployment spells, relationships with the agency, and the extent of usage of the institution's digital services (see Appendix A, Table 7).

3.4.4 Measuring Take-up

To effectively optimize this strategy throughout the experiment, we need a measurable outcome that reflects the email's impact. When assessing the success of an email, we have the option to consider various outcomes. These outcomes can range from very short-term actions like opening an email, clicking on an email link and navigating on the website, to medium-term actions such as updating

[7]We constitute a large control group since observing individuals in this group has no cost and the information is already available in the administrative data. This allows us to ensure statistical power.

and publishing the skills profile, increasing visibility of the profile. We think that these short term outcomes and midterm outcomes are correlated. Our theory of change is illustrated in figure 4.

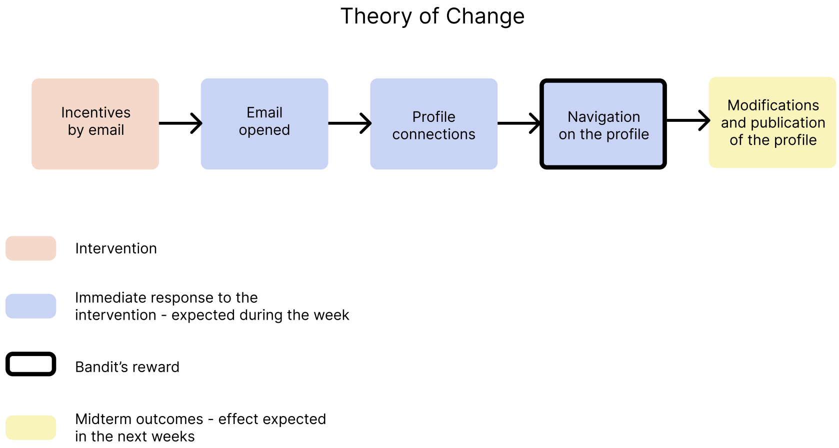


Figure 4: Expected consequences of our intervention - short and mid terms

There is a trade-off to consider before choosing one of these outcomes. Short-term outcomes offer the advantage of providing rapid feedback, allowing us to concentrate our experimentation over a relatively short time frame. This is all the more important as, over a long period of time, the skills profile can be improved over time by its developers, and we want to avoid any kind of exogenous shock of this type. Conversely, longer-term outcomes align more closely with the implementers' interests, that is to encourage job seekers to improve their profile, fill it progressively or publish it. However, such outcomes may be more diluted and impact a smaller segment of the population, depending on our measure of a "filled profile", making them less suitable for fine-grained optimization during the experiment.

As a result, we have opted for a short-term outcome that provides rapid feedback.

$$Y_i^{\text{obs}} = \begin{cases} 1 & \text{if the jobseeker navigates to the profile webpages within one week} \\ 0 & \text{otherwise.} \end{cases}$$

In our approach, an email is considered to have a positive impact if the jobseeker accesses their PES account and navigates to the profile web pages within one week of receiving the email. Although this short-term outcome may not appear directly linked to the final quality of the profile, it is reasonable to assume a significant correlation between this behavior and other indicators of website activity and profile quality. Navigating the skills profile represents an initial and essential step, indicating that the email has been opened, the individual has logged into the platform, and the individual has engaged with their profile. Consequently, we believe that this short-term outcome best serves the objectives of the bandit algorithm and aligns with the design of our experiment. Moreover, all our incentives have been designed to specifically engage more the jobseekers with the platform, and it is therefore quite natural to measure their activity on the platform as a sign of success of the emails.

4 Overall Impact of the Experiment on Platform Use

4.1 Assessing the Global Impact of the Mailing Campaign

To evaluate the impact of the global mailing campaign on profile usage, we compare two groups : the control group; which did not receive any email, and the treated group, which received one. A balance-check provided in Appendix D shows that covariates are balanced between the two randomly sampled groups. Since assignment to the control and treated groups is randomized as in a standard randomized control trial, naive intent-to-treat estimators apply without complications when assessing the experiment's impact. However, treatment should be interpreted as being assigned to the treatment group, i.e receiving an email according to the algorithmic policy. We estimate treatment impact using an ordinary least square regression with batch fixed effects.

$$Y_i^{obs} = \beta_0 + \beta_1 T_i + \sum_{b=2}^{B} \alpha_b \mathbb{1}\{b_i = b\} + \epsilon_i \tag{1}$$

where Y_i^{obs} corresponds to whether i took up the profile, T_i is a dummy indicating whether i was assigned to the treated group, and b_i denotes the batch during which i was assigned to treatment or control group. This assessment is conducted on the 252,000 individuals sampled during the main experiment. While we focus on log-ins to profile-related web pages as a main outcome, we run the same specification on other measures of profile use in the next section.

4.2 Impact of Mailing Campaign on Take-up Measures

Table 1 presents reports the results of the intent-to-treat regression, as described in equation 1 for various outcome measures. Column 1 and 2 show treatment effect for the main outcome of interest, the probability to visit the profile. Additionally, Columns 3-8 provide treatment effects for other proxies of user engagement with the profile. These treatment effects are evaluated at two different time windows, namely 7 and 15 days following the email sending. The first graph of Figure 5 displays the email treatment effect across all time windows, ranging from 0 to 30 days.

Main outcome Treatment increases significantly the share of participants who visit the profile page in a 7-day time window after receiving the email. On average, being sent an email results in a 2.46 percentage point increase in the likelihood of logging in during this 7-day period. Despite the relatively low compliance rate, the impact on traffic is substantial. The email intervention leads to an increase of 90 % increase in traffic with 5.3% of email recipients logging into the platform compared to 2.8% in the control group. However, the compliance rate is relatively low in absolute term. Treatment effect on the main outcome remains constant and significant after 15 days. We selected the 7-day time window as a reference period for the update of the adaptive algorithm. Figure 5 visually represents the email's impact on user behavior and shows that it occurs mainly within the first week.

Other proxies of profile usage The chosen outcome driving the algorithm may only partially reflect user engagement with the profile. To address this concern and verify the extent of user engagement beyond the initial visit to the profile webpage, we present treatment effects on several proxies of engagement. Firstly, we examine the treatment effects on the number of visits (Column 3-4 of Table 1), which provides insight into user fidelity and the frequency of engagement. This measure gives for each individual, the number of days she visits the profile page. The treatment effect is statistically significant at 7 days and further increases at 15 days. This indicates that individuals who visit the profile page also continue to engage with it after one week. Additionally, we explore the treatment effects on two other proxies : the probability of modifying at least on element of the profile and the probability of publishing or refreshing[8] the profile on the platform. These measures provide a sense of user effort in improving their profile and making it visible to recruiters. The treatment effect is statistically significant at 7 and 15 days for both the likelihood to modify and to publish the profile. However, the magnitudes of these effects are lower compared to the probability of visiting the profile. On average, being sent an email results in a 1.6 percentage point increase in the likelihood of modifying the profile during the 7-day period and a 0.04 percentage point in publishing the profile. It represents respectively a 72% increase and a 62% increase at 7 days.

	Any visit (Main)		# of visits		Any modification		Any publication	
	At 7 days	At 15 days	At 7 days	At 15 days	At 7 days	At 15 days	At 7 days	At 15 days
Treatment	0.025***	0.024***	0.025***	0.030***	0.016***	0.016***	0.005***	0.005***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.000)	(0.000)
Control Mean	0.028	0.057	0.031	0.075	0.022	0.039	0.008	0.015
Observations	252,000	252,000	252,000	252,000	252,000	252,000	252,000	252,000

Note: *p<0.1; **p<0.05; ***p<0.01

This table shows results from equation (1) for different outcomes of interest observed in a 7 and 15 days window after sending the email: visiting at least once the skills profile, the number of days during which the skills profile has been visited and the probability of modifying at least one element of the skills profile and the probability to publish the profile on the platform (or to update the publication). All regressions include batches fixed effects. Standard errors are clustered at the batches level.

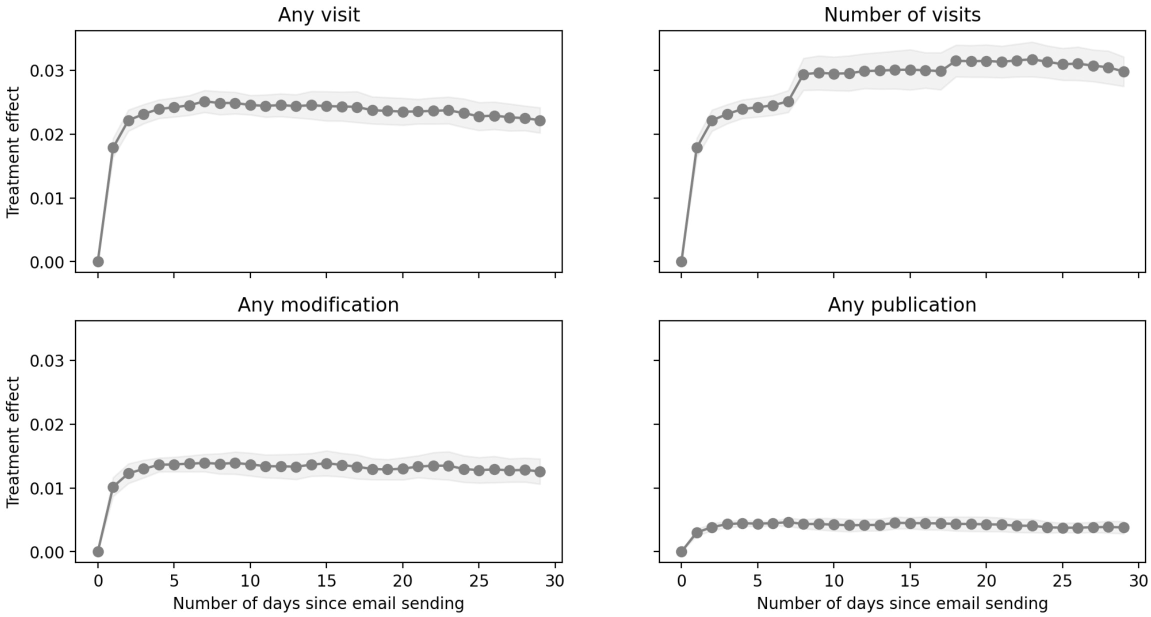
Table 1: Treatment Effects on profile usage

Overall, sending an email had a positive impact on jobseekers' engagement with their profiles. Although the take-up rate of the email is relatively low, it still results in a substantial increase compared to the control group. The email not only prompted visits to the profile page but also drove jobseekers to explore their profiles further and take active actions through modifications and publications of the profiles, albeit with a lower magnitude.

4.3 Heterogeneity in Response Probabilities

Although not all jobseekers take up with the email, a significant portion do respond positively by visiting the profile page. In this section we attempt to describe the characteristics of jobseekers who exhibit the highest and lowest compliance rate as a result of email sending. We use a data-driven method developed by Chernozhukov et al. (2019) called the generic machine learning method. It strives to leverage machine learning methods to explore heterogeneity, while relying on sample splitting to provide rigorous statistical inference. To explore heterogeneity in treatment effects, the Generic Machine Learning approach suggests the following procedure. First, machine learning methods are used to create a proxy predictor for the Conditional Average Treatment Effect (CATE) using a subset of the data (auxiliary sample), that is the effect of being assigned to the treated group

[8] If individuals already have a published profile they can refresh it in one click in order to signal they are still interested in being on the platform, we count this action as a publication



Notes : This figure shows estimated coefficients for β_1 and corresponding confidence intervals (95%) in equation 1 computed day by day for a month (30 days). Four profile usage proxies are presented, one in each square subfigure. The regression includes fixed effects for batches and standard errors are clustered at the batches level. The estimation is conducted on 252 000 individuals from the main experiment.

Figure 5: Profile usage over time

instead of the control group, conditional on the characteristics X_i. In the second stage, the estimated function is used to make predictions for the remaining observations (main sample, referred to later as M). These predictions are then used to perform inference on three specific moments of the CATE :

- Best linear predictor (BLP) : It provides an estimate of the Average Treatment Effect (ATE) and a Heterogeneity Loading (HET) parameter, which assesses the presence of heterogeneity in treatment effects.

$$\begin{aligned} Y_i =& \theta_1 \hat{B}(X_i) + \theta_2 \widehat{\mathrm{CATE}}(X_i) + \beta_{\mathrm{ATE}}\big(T_i - p(X_i)\big) \\ &+ \beta_{\mathrm{HET}}\big(T_i - p(X_i)\big)\Big(\widehat{\mathrm{CATE}}(X_i) - \overline{\mathrm{CATE}}_M\Big) + \delta_i + \epsilon_i, \end{aligned} \tag{2}$$

  where $p(X_i)$ is the re-estimated probability of being assigned to the treatment group conditional on X_i, $\hat{B}(X_i)$ is an estimation of the baseline potential outcome conditional on X_i based on the observed outcomes in the control group, $\widehat{\mathrm{CATE}}(X_i)$ is an estimate of the conditional average treatment effect obtained through different machine learning techniques, $\overline{\mathrm{CATE}}_M = \frac{1}{\#\{i \in M\}} \sum_{i \in M} \widehat{\mathrm{CATE}}(X_i)$ is the mean predicted impact on sample M used for the regressions, $\delta_i = \sum_{b=1}^{B} \alpha_b \mathbb{1}\{b_i = b\}$ represents the fixed effects, and we use weights $w(X_i) = \frac{1}{p(X_i)(1-p(X_i))}$. β_{ATE} measures the mean treatment effect. The coefficient of interest is β_{HET}: it is different from zero if $\widehat{\mathrm{CATE}}(X_i)$ captured heterogeneous effects.

- Group Average Treatment Effects (GATES): Observations are grouped into subgroups based on their predicted treatment effects. This allows for the examination of treatment effects within these specific groups.

$$Y_i = \theta_0 + \theta_1 \hat{B}(X_i) + \theta_2 \widehat{\text{CATE}}(X_i) + \sum_{k=1}^{5} G_k \cdot (T_i - p(X_i)) \cdot Q_{ki} + \delta_i + \epsilon_i \tag{3}$$

To ensure robustness, this procedure is repeated on multiple sample splits. The estimated parameters are then aggregated by taking the median values and adjusting pvalues to take into account splitting uncertainty.

We use all jobseekers' characteristics presented in Table 7 and the main outcome of interest which is the visit to the profile page within a 7 day time window.

Best Linear Predictor Table 2 provides estimates of the coefficients β_{ATE} and β_{HET}, which correspond to the Average Treatment Effect (ATE) and Heterogeneity Loading (HET) parameters in the Best Linear Predictor. We provide estimates for Elastic Net which outperformed other method in detecting heterogeneity and give estimates for less effective machine learning methods as a robustness check in Appendix F.3. Reassuringly we obtain the same estimate for the average treatment effect as observed in Column 1 of Table 1 and sending an email leads to a significant increase of 2.5 percentage points in the proportion of individuals visiting the profile. Additionally, we find strong evidence of heterogeneity in treatment effects, as indicated by the statistically significant estimate of β_{HET}. This suggests that the impact of the email varies across different segments of the population. The estimates of β_{HET} are close to 1, indicating that the machine learning proxies employed are good predictors of the conditional average treatment effect.

	β_{ATE}	β_{HET}
Estimate	0.0244	1.31
Conf. interv.	[0.0218, 0.0272]	[1.02, 1.56]
Corrected pvalue	0.000	0.000

Table 2: Average treatment effect and detection of heterogeneous treatment effects

Notes: Sample size = 252,000 individuals. These results are estimations from equation [2] (in 4.3). Following Chernozhukov et al. (2019), all displayed estimations are medians over 20 splits in half, confidence intervals are adjusted, and so are p-values. Standard errors are clustered at the batch level.

	Least affected (G_1)	Most affected (G_5)	Difference (G_5- G_1)
Estimate	0.00923	0.0538	0.0449
Conf. interv.	[0.00427 , 0.0143]	[0.0467, 0.0602]	[0.0372, 0.0527]
Corrected pvalue	0.007	0.000	0.000

Table 3: GATES estimates for least and most affected groups

Notes: Sample size = 252,000 individuals. Estimates for G_5 and G_1 rely on equation [3] (in 4.3). We also test whether the difference between both is significantly different from 0. Following Chernozhukov et al. (2019), all displayed estimations are medians over 20 splits in half, confidence intervals are adjusted, and so are p-values. Standard errors are clustered at the batch level.

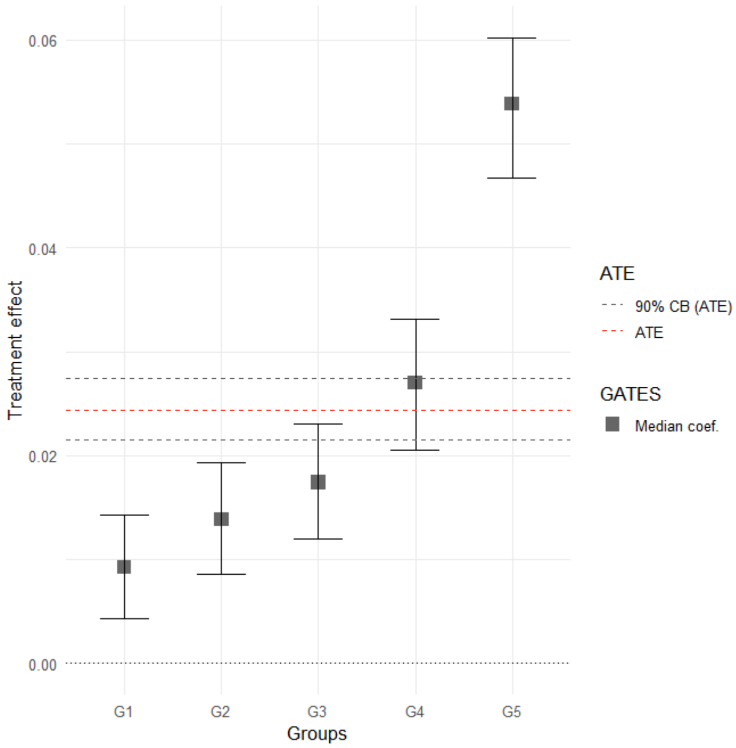


Figure 6: GATES

Notes: Sample size = 252,000 individuals. Number of repetitions: 20. We report GATES estimates for the honest generalized random forest. The group heterogeneity score (x-axis) corresponds to an estimated quintile of the ITT. 90% confidence bounds are computed following Chernozhukov et al. (2019). Standard errors are clustered at the batch level.

Group Average Treatment Effect Next, we estimate the Group Average Treatment Effects (GATES) by quintiles of the machine learning proxies. The results of the hypothesis test, presented in Table 3, confirm that the difference in Average Treatment Effects (ATE) between the most and least affected groups amounts to 0.042 and is statistically significant. Figure 6 provides a visual representation of the estimated GATES coefficients along with joint confidence intervals. Reassuringly all groups show positive and significant treatment effect. However, it appears clearly that the most affected group (G_5) demonstrates really much higher treatment effect to the email compared to the other groups. For this subgroup, representing the top 20% of the sample, the email increased user engagement by 5.2 percentage point compared to 1 percentage point on average in the bottom 20%.

Classification Analysis Finally, we turn to the analysis of the characteristics of the most and least affected groups (i.e the Classification analysis, or CLAN estimates in Chernozhukov et al. (2019)'s terminology). Table 4 reports the Classification Analysis (CLAN) estimates for a selected set of covariates. These estimates are derived by regressing dummies for each characteristic, with fixed effects corresponding to the batches. To simplify interpretation, continuous variables have been dichotomized. Additional covariate estimates can be found in Appendix 10. The estimated differences in means between the most and least affected groups are found to be statistically significant for the majority of the covariates. Jobseekers belonging to the most affected group have more experience in their job, have higher reservation wages, and are more likely to belong to the executive qualification level. This suggests that jobseekers who are already relatively close to employment or less vulnerable are more likely to engage in their profile after receiving the email. One possible interpretation is that they are already familiar with digital job search processes which makes them more likely to be aware of job platforms and their effectiveness. Additionally, jobseekers who have recently started looking for a job are more likely to visit their profile page after receiving the email. This finding suggests that individuals who have recently started looking for a job are more motivated to engage with their profiles.

Furthermore, individuals who already have some level of proximity to the institution, such as having had contacts within the previous 3 months or having visited the Pôle emploi's website in the last month, are overrepresented among the group that has navigated to their profile page following the intervention. This indicates that those who already have some level of engagement with the employment institution are more likely to be responsive to the email. This may mean that a simple email alone may not be sufficient to engage jobseekers who are more distant from the institution, who might need a stronger pedagogy or a more in-depth support.

	20% Least	20% Most	Difference	pvalue
Experience				
Below 5 years	0.79	0.38	-0.41	0.00
Between 5 and 10 years	0.13	0.14	0.00	0.00
Above 10 years	0.08	0.47	0.40	0.00
Reservation Wage				
Below 25	0.26	0.19	-0.08	0.00
Between 25 and 50	0.23	0.15	-0.08	0.00
Between 50 and 75	0.23	0.24	0.01	0.00
Above 75	0.28	0.42	0.12	0.00
Qualification Level				
Skilled Worker	0.15	0.13	-0.02	0.00
Non qualified employee	0.20	0.16	-0.05	0.00
Qualified employee	0.53	0.50	-0.03	0.00
Executive	0.12	0.22	0.10	0.00
Age				
Below 25	0.17	0.03	-0.14	0.00
Between 25 and 39	0.61	0.17	-0.45	0.00
Between 40 and 54	0.17	0.36	0.19	0.00
Above 55	0.00	0.42	0.42	0.00
Gender				
Is male	0.52	0.48	-0.03	0.00
Relationship with institution				
Already published a profile	0.21	0.21	-0.00	0.02
At least one visit in the last month	0.11	0.67	0.56	0.00
At least one meeting in the last month	0.17	0.53	0.36	0.00
Number of days since last registration				
Below 3 months	0.07	0.29	0.22	0.00
Between 3 months and 1 year	0.93	0.71	-0.22	0.00
Between 1 year and 2 years	0.56	0.39	-0.17	0.00
Above 2 years	0.27	0.24	-0.02	0.00

Table 4: Classification analysis

Note: The following table presents the average characteristics of the groups classified as the least and most affected. The table also includes the results of systematic tests conducted to examine the differences between these two groups.

Altogether, the overall impact of providing jobseekers with incentives by email to fill in the profile on their behavior. Jobseekers that have received an email responded by going more to their online profile compared to the control group. They also undertook active actions by modifying it or publishing their profile on the platform. However the global take-up rate is relatively low. A deeper analysis of heterogeneity reveals that a subgroup representing 20% of our sample reacted much more than other subgroups and drives our results. This subgroup consists mainly in jobseekers were already close to employment and had pre-existing proximity with the institution.

5 Algorithm Behavior Throughout the Experiment and Variants Impact

5.1 Effectiveness of the contextual bandit strategy in the main experiment

We examine whether the use of the contextual bandit strategy, with its potential to cast aside lackluster e-mails and to leverage heterogeneity, increased the platform's take-up rate in the main experiment.

The consideration of within-batch treatment effects, depicted in Figures 7 and 8 alongside 95% confidence intervals, is sobering in that regard. If a better policy than uniform randomization of e-mail construction had been learned, treatment effects would have increased throughout time with respect to the first batch (which had been sent uniformly at random). Since the first batch is the only batch in which treatments are allocated completely at random, we will use the treatment group from this batch as a reference for assessing the effect of random email allocation. Therefore, we propose to test the impact of belonging to one of the batches during which the allocation of treatment arms is no longer random but is determined by the algorithm[9]. However, no clear increase of treatment effects seems to have occurred. To test this formally, we also consider the specification:

$$Y_i^{obs} = \beta_0 + \beta_1 T_i + \beta_2 1\{b_i \neq 1\} T_i + \sum_{b=2}^{B} \alpha_b \mathbb{1}\{b_i = b\} + \epsilon_i \tag{4}$$

If using a contextual bandit strategy had resulted in an improvement over random choice of sent emails, one would expect that $\beta_2 > 0$; yet the estimate for β_2 in this setup is -0.0003 ($p = 0.907$).

[9] We could have tested alternative specifications by considering only the latest batches if we had thought that the initial batches were still part of a learning phase. However, we prefer to use only the first batch as a reference because we believe that it already contains a large number of individuals, and we want to ensure that the reference group is allocated fully at random. Additionally, we are also suggesting a test of an optimal policy trained off-policy to assess if more can be learned from a large sample.

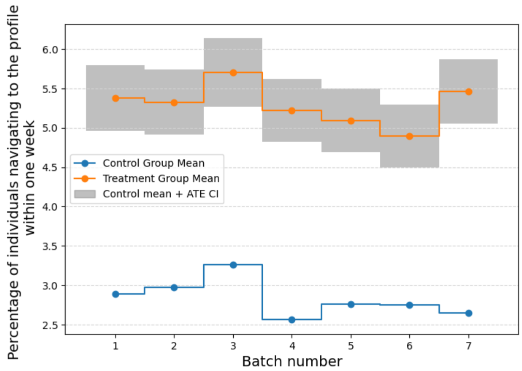


Figure 7: Evolution of the control and treatment means across time

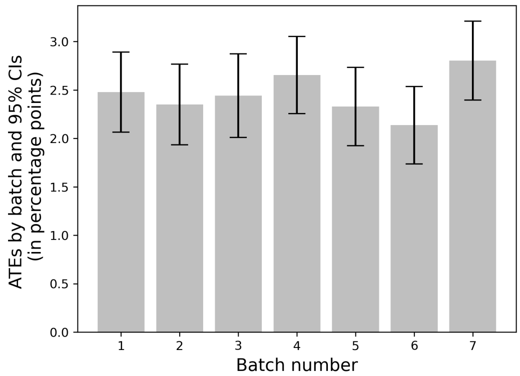


Figure 8: Evolution of treatment effects across time

The figure on the left shows the control and treatment means for each batch. The figure on the right shows the treatment effects at each batch, comparing the treated group and the control group. For each bar, 12 000 treated individuals are compared with 24 000 non treated individuals observed simultaneously. The treatment effect is estimated using a linear regression of the observed outcome on the treatment variable separately for each batch.

5.2 Assessing and Evaluating Behavioral and Informational Levers

5.2.1 Identification strategy

At a high level, we wish to measure whether including lever l in an email (*e.g.* "including information on the number of recruiters using the platform") increased take-up.

An issue arises in defining the notion of impact that should be studied: including lever l in an otherwise empty email may not have the same impact as adding l to a long email where all other levers l' (compatible with the presence of lever l) are present. To solve this issue, we integrate over all possible combinations of levers l' (giving each possible email equal weight) to define the quantity of interest. More precisely, for a given lever l, let us define communication policies:

$$\pi_l(w) := \frac{\mathbb{1}\{w \text{ contains } l\}}{\sum_{w'=1}^{32} \mathbb{1}\{w' \text{ contains } l\}}$$

$$\pi_{\bar{l}}(w) := \frac{\mathbb{1}\{w \text{ does not contain } l\}}{\sum_{w'=1}^{32} \mathbb{1}\{w' \text{ does not contain } l\}}$$

In other word, π_l sends emails at random among all email structures w containing lever l, whereas $\pi_{\bar{l}}$ sends emails at random among email structures *not* containing lever l.

Adopting Rubin and Neyman's potential outcome framework, and denoting $Y_i(w)$ the potential outcome of individual i under email w, the goal is to estimate the quantity:

$$\Delta_l = \mathbb{E}\left[\sum_{w=1}^{K} \pi_l(w) Y_i(w)\right] - \mathbb{E}\left[\sum_{w=1}^{K} \pi_{\bar{l}}(w) Y_i(w)\right] \tag{5}$$

which measures the average difference in outcomes when individuals are assigned treatments according to π_l and to $\pi_{\bar{l}}$.

As highlighted for instance by Hadad et al. (2021), the study of adaptively collected data is more involved than in a standard randomized control trial. Since treatment assignment probability for individual i depends on his or her context, as well as on previously collected data used as input to the assignment algorithm, observations $\{(Y_i^{obs}, W_i, X_i)\}_{i=1\ldots N}$ are no longer independent and identically distributed.

In the present work, we restrict our consideration to the Augmented Inverse Propensity Weighting (AIPW) estimator:

$$\hat{\Delta}_l = \frac{1}{N}\sum_{i=1}^{N}\hat{\Gamma}_i(X_i, \pi_l - \pi_{\bar{l}})$$

$$\hat{\Gamma}_i(X_i, \pi_l - \pi_{\bar{l}}) = \sum_{w=1}^{W}\big(\pi_l(w) - \pi_{\bar{l}}(w)\big)\left(\hat{\mu}_i(X_i, w) + \frac{\mathbb{1}\{W_i = w\}}{e_i(X_i, w)}\big(Y_i^{\text{obs}} - \hat{\mu}_i(X_i, w)\big)\right)$$

Here, $\hat{\mu}$ denotes an estimator predicting potential outcome $Y_i(w)$ of individual i under treatment w using data collected before individual i's assignment. In practice, $\hat{\mu}$ is instantiated as a causal forest trained on all the characteristics listed in table 7, with default hyper-parameters, and on a rolling 4,000-by-4,000 basis. This estimator is unbiased [10], but not necessarily asymptotically Gaussian, prohibiting the formation of valid confidence intervals. [11]

This assessment is conducted on the 84,000 treated individuals sampled during the main experiment.

5.2.2 Impact Estimations of Email Variants on Take-up

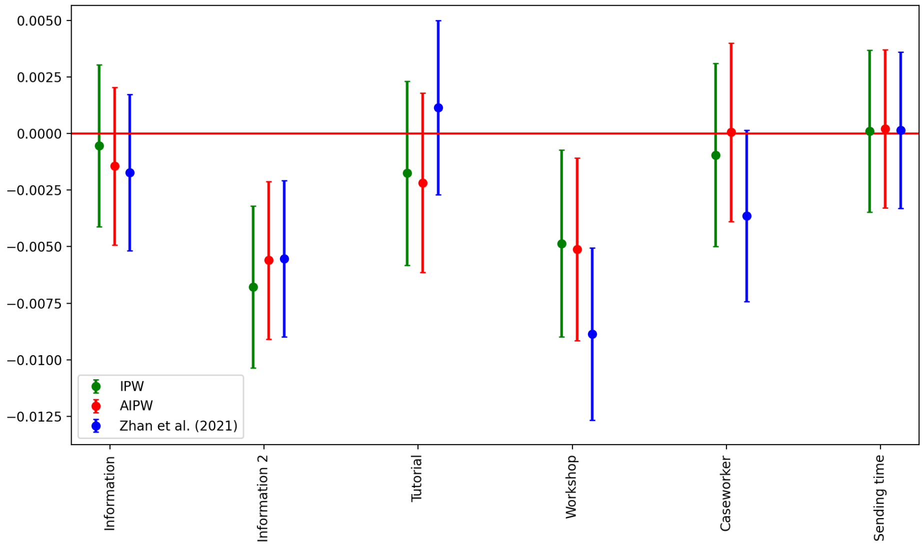
Results for the impact of each behavioral lever using Augmented Inverse Propensity Weighing estimators are presented in Table 5. Results for two other similar estimators - namely Inverse Propensity Weighting (i.e. AIPW without augmentation using a model $\hat{\mu}_i(X_i, w)$) and contextually weighted estimators proposed by Zhan et al. (2021) (which amount to re-weighting AIPW scores to reduce variance, allowing the formation of valid confidence intervals under further assumptions) - are also presented as a robustness check.

Recall that these estimators assess the impact of adding a behavioral lever l to an email, integrating over the inclusion or not of all other possible communication variants l'. The nullity of the impact of behavior levers consisting in the provision of information, of a tutorial, or in the variation of sending time cannot be rejected. All three considered estimators agree in finding a negative impact of adding information about improvements in service quality, and of mentioning a profile-related workshop. The inclusion of a call-to-action button to schedule an appointment with a caseworker appears to have little, or a slightly negative, impact.

[10] See *e.g.* Zhan et al. (2021), appendix A.1.

[11] An avenue for further work would involve the use of re-weighted estimators, such as those proposed by Zhan et al. (2021) or Bibaut et al. (2021), for which confidence intervals can be formed under further assumptions on the logging policy or estimators $\hat{\mu}$.

Figure 9: Estimators for the impact of communication variants



This figure shows the results for the estimation of Δ_L in equation (5) using various strategies: IPW estimation using the sampling weights, AIPW using the sampling weights and estimating the outcome models using a causal forest, and also report estimates following Zhan et al. (2021) as a robustness check. The estimation is conducted on the 84,000 treated individuals in the adaptive experiment. Detailed numeric results are provided in table 5.

Table 5: Estimators for the impact of communication variants

	Information 1	Information 2	Tutorial	Workshop	Caseworker	Sending time
AIPW	-0.002	-0.006	-0.002	-0.005	-0.0	-0.0
AIPW CI	[-0.005, 0.002]	[-0.009, -0.002]	[-0.006, 0.002]	[-0.009, -0.001]	[-0.004, 0.004]	[-0.004, 0.004]
AIPW SE	0.002	0.002	0.002	0.002	0.002	0.002
IPW	-0.001	-0.007	-0.002	-0.005	-0.001	-0.0
IPW CI	[-0.004, 0.003]	[-0.011, -0.003]	[-0.006, 0.002]	[-0.009, -0.001]	[-0.005, 0.003]	[-0.004, 0.003]
IPW SE	0.002	0.002	0.002	0.002	0.002	0.002
Zhan Mean	-0.002	-0.006	0.001	-0.009	-0.004	-0.0
Zhan CI	[-0.005, 0.002]	[-0.009, -0.002]	[-0.003, 0.005]	[-0.013, -0.005]	[-0.008, -0.0]	[-0.004, 0.003]
Zhan SE	0.002	0.002	0.002	0.002	0.002	0.002

This table shows the results for the estimation of Δ_L in equation (5) using various strategies: IPW estimation using the sampling weights, AIPW using the sampling weights and estimating the outcome models using a causal forest, and also report estimates following Zhan et al. (2021) as a robustness check. The estimation is conducted on the 84,000 treated individuals in the adaptive experiment. These results are represented graphically in figure 9.

6 Learning and Assessing an Optimized Allocation Policy

There may be several reasons explaining why no interventions were highlighted during the data collection. Firstly, one possible reason could be that we allowed for too fine grained exploration of heterogeneity. While the use of a non parametric estimator for the conditional mean enables us to capture nuanced treatment effects and account for individual differences, it may have impeded rapid convergence towards the treatments that work best on average. The other reason may lie in the design of our experiment, i.e. low-cost e-mail interventions are not sufficiently different to leverage on average and by subgroups a significant better allocation strategy. In this section, we propose to use the data collected in order to estimate an optimal personalized allocation strategy of emails and to compare its performance with a policy that would allocate emails at random.

6.1 Hypothesis and Design

We want to test whether learning from collected data about what email works for whom enables us to define a strategy that works best than random email assignment. The two assignment strategies (policies) are characterized as mappings π from $\mathcal{X} \times \mathcal{W}$ to $[0,1]$ that are probabilities (i.e. such that for all $x \in \mathcal{X}$, $\sum_{w \in \mathcal{W}} \pi(x,w) = 1$). Let us denote π^r an allocation policy that assigns treatment arms to jobseekers at random; that is,

$$\forall (x,w), \quad \pi^r(x,w) = \frac{1}{|\mathcal{W}|}$$

Let us also denote $\hat{\pi}^c$ an optimized contextual allocation policy learned from data. We want to test the following statement :

$$\mathcal{H}_0 : \mathbb{E}[\sum_{w \in \mathcal{W}} \hat{\pi}^c(X_i,w) Y_i(w)] = \mathbb{E}[\sum_{w \in \mathcal{W}} \pi^r(X_i,w) Y_i(w)]$$

$$\mathcal{H}_a : \mathbb{E}[\sum_{w \in \mathcal{W}} \hat{\pi}^c(X_i,w) Y_i(w)] \neq \mathbb{E}[\sum_{w \in \mathcal{W}} \pi^r(X_i,w) Y_i(w)]$$

The null hypothesis states that the contextual allocation policy ($\hat{\pi}^c$) performs as well as the random allocation policy (π^r) while the alternative hypothesis states that the contextual allocation policy ($\hat{\pi}^c$) gives a higher reward than the random allocation policy (π^r).

In order to test this hypothesis we use the following design. 228,000 individuals are randomly sampled from the eligible population of jobseekers. Just as in the main experiment, the eligible population is divided into two groups: the control group and the treatment group. The division is done randomly, with a probability of 2/3 for the control group and 1/3 for the treatment group. Within the treatment group, individuals are randomly assigned, with an equal probability, to either the optimal allocation policy ($\hat{\pi}^c$) or the random allocation policy (π^r). Each group receives emails according to their assigned allocation policy. The control group follows the random allocation policy (π^r), while the treatment group follows the contextual allocation policy learned from data ($\hat{\pi}^c$). We finally compare outcomes of the group treated with the random policy and of the group treated with the contextual policy.

6.2 Constructing an Optimized Allocation Policy

This section describes the construction of the contextual optimal policy $\hat{\pi}^c(x_i,.)$. It will be deterministic, that is, for all x and w, $\hat{\pi}^c(x,w) \in \{0,1\}$.

Compute doubly robust scores estimation Using data collected adaptively in the first phrase, we first compute doubly robust scores:

$$\hat{\Gamma}_i(X_i) = \left(\hat{\mu}_i(X_i,w) + \frac{\mathbb{1}\{W_i = w\}}{e_i(X_i,w)} \left(Y_i^{\text{obs}} - \hat{\mu}_i(X_i,w)\right) \right)$$

where $\hat{\mu}_i(X_i,w)$ is an estimate for the conditional mean for email w and context X_i (built only using observations gathered before or during the batch i was assigned a treatment [12]), and $e_i(X_i,w)$ is the

[12]This standard practice in the literature (justified by the time dependence in the data collection process) enables the preservation of the unbiased character of the doubly robust scores.

assignment probability of the individual to arm w (according to the assignment policy used for i's batch in the first phase).

Assignment probabilities $e_i(X_i, w)$ are known due to the use of a Thompson sampling procedure, and are plugged directly in the computation of the doubly robust scores.

Conditional mean estimates $\hat{\mu}_i(X_i, w)$ are built in the following fashion. Individual indexes are shuffled and sorted by batch. $\hat{\mu}$ is instantiated as a causal forest, with default hyper-parameters (including the honesty property), and trained on a rolling 4,000-by-4,000 basis. That is, estimate $\hat{\mu}_i(X_i, w)$ for individual i is obtained from a model trained using observations 1, $\ldots \lfloor \frac{i}{4000} \rfloor \times 4000$ (when $i < 4{,}000$, predictions are initialized to zero). Finally, AIPW estimates are clipped in order to promote stability (Su et al. 2020). A pseudo-code describing the estimation procedure is provided in Appendix H.

Selecting most promising emails and relevant heterogeneity The following preprocessing procedure, advocated by Athey et al. (2022) is applied before computing the optimal contextual policy.

1. We drop arms that are less promising based on the frequency of assignment of emails during the data collection phase. For each arm we count how often they have been assigned during the first phase.

$$\text{freq}(w) = \frac{1}{N} \sum_{i=1}^{n} \mathbb{1}\{W_i = w\}$$

   After sorting in descending order emails according to the frequency of assignment, we keep the top half of them as valid candidate observations for the contextual policy and drop individuals that were not assigned those emails.

2. We restrict the set of covariates allowed to serve for the estimation of the contextual policy. To do so, we train a causal forest on the data and retain only the top 50 % covariates ranked by feature importance. This first selection of the covariates prevents us from splitting by mistake on an irrelevant covariate that only plays a role when looking at a very fine level of heterogeneity while letting more important ones aside A second advantage is that training a policy tree over a large number of covariates is computationally demanding.

Optimal policy estimation The last step involves estimating a tree policy following Sverdrup et al. (2020). The method seeks to find a simple rule-based policy for treatment allocation after the adaptive experiment (off-policy) by fitting a shallow but globally optimal decision tree (in contrast to usual greedy tree-building strategies). The depth of the tree is chosen through cross-validation by fitting a tree on the first 75 % observations and evaluating on the remaining data. We select the depth that leads to highest take-up rate based on doubly robust scores. A pseudo-code describing the procedure of the policy-tree construction is provided in Appendix H.

Figure 10: Allocation decision based on the estimated policy tree

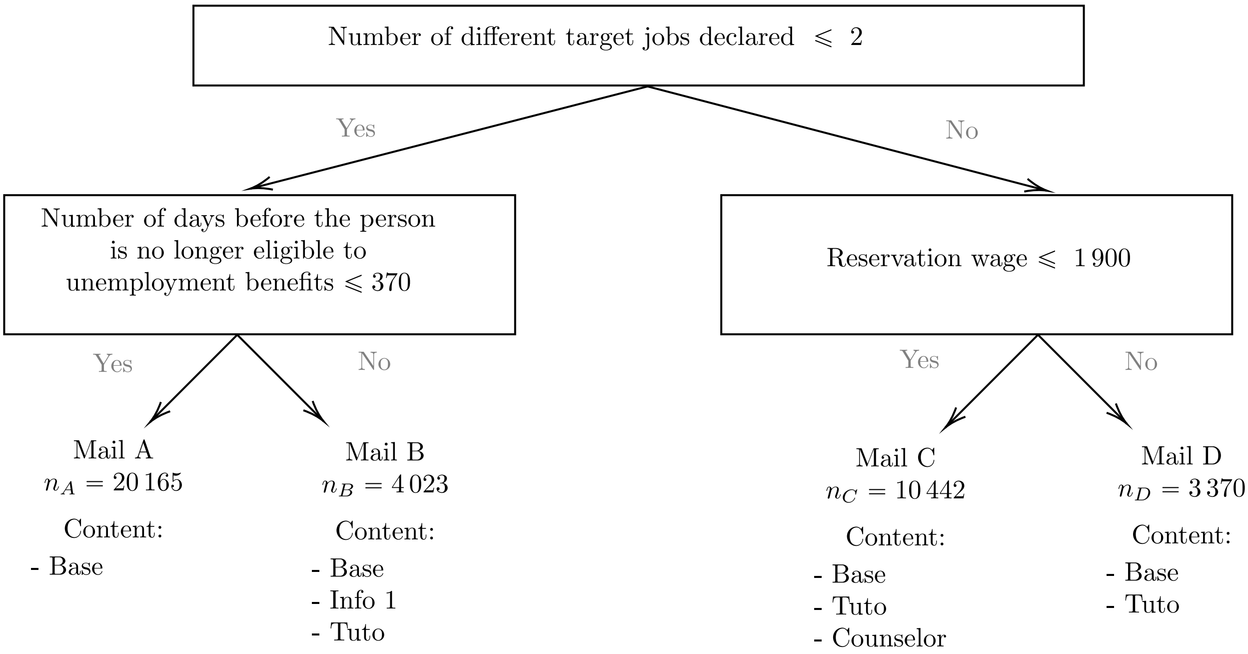


Figure 10 illustrates the deterministic allocation rule or policy derived from the policy learning algorithm, which maps jobseeker features to our different treatments. The learned policy aligns with some of our primary hypotheses regarding heterogeneity in treatment effects.

According to the learned policy, jobseekers who have fewer than two targeted jobs and have been unemployed for less than a year should receive the minimal email intervention. On the other hand, those who have been unemployed for a longer duration should receive additional information about the number of recruiters on the platform and a basic list of steps to follow. Furthermore, individuals who have declared searching in more than two sectors and have a relatively high reservation wage should be provided with a tutorial, while those with a lower reservation wage should be offered human assistance through an appointment with a counselor. Considering that those declaring a lower reservation wage are also the less qualified, this policy is consistent with the intuition that individuals with lower qualifications may benefit the most from human support.

6.3 Impact of Personalization on Take-up

We next turn to the evaluation of the magnitude of the gains associated with the use of an individualized communication strategy. Jobseekers were randomly assigned either to the personalized allocation ($\hat{\pi}^c$), the random allocation (π^r) or the control group (no emails sent). We measure the impact of having been assigned to the random and the personalized allocation by running the following regression:

$$Y_i^{\text{obs}} = \gamma_0 + \gamma_1 T_i + \gamma_2 (T_i \times P_i) + u_i \tag{6}$$

where T_i is a dummy for having been assigned to the treatment group and P_i is a dummy indicating that the individual has been assigned to the personalized allocation group[13]. γ_0 represents the log-in probability in the control group. γ_1 gives the benefit from being assigned to the treatment group. γ_2 can be interpreted as the additional gains from being assigned to the personalized group.

	Any visit
	At 7 days
Intercept	0.028***
	(0.000)
Treatment	0.022***
	(0.001)
Optimized Allocation	0.002*
	(0.001)
Observations	228,000

Notes: *p<0.1; **p<0.05; ***p<0.01

This table show results from equation (6) on 228, 000 individuals, 2/3 of which do not receive any email and 76, 000 are randomly assigned or to the random policy or to the personalized policy.

Table 6: Personalized allocation vs Random allocation

The regression results comparing the two policies using the run-off batch data are presented in Table 6. Regardless of the policy assigned, receiving any email increases the share of jobseekers who visit their profile by 2.2 percentage point compared to the control group. While jobseekers treated by the optimized contextual policy logged slightly more often in a profile-related web page (by 0.2 percentage points) than the recipients of the uniformly randomized policy, this difference is not statistically significant at 5% ($p = 0.07$). These findings suggest that the personalized allocation strategy (as depicted in Figure 10), based on the data collected during the first phase, did not increase take-up rates compared to a random allocation strategy. We conclude that e-mails are not sufficiently different for a personalized strategy to improve significantly take-up rates.

7 Discussion

In this section, we would like to examine the specifics of adaptive design, as its use in public policy evaluation is relatively new.

Finding the right balance between exploration and exploitation Finding the optimal balance between exploration and exploitation is a critical aspect of adaptive experiment design. Several design choices impact the explore-exploit balance, including:

- Specification choice : The use of a non-parametric estimator, which avoids making assumptions about the relationship between covariates and the outcome variable, can be advantageous in capturing heterogeneity but may hinder the speed of exploitation.

[13]Note that in this regression, the interaction between T_i and P_i is redundant, since T_i is always 1 when P_i is 1. The only purpose of the interaction here is to clarify this relationship and remove any doubt on this issue.

- Minimum probability threshold : Setting a high minimum probability threshold helps prevent premature exploitation of suboptimal treatment arms. It also maintains assignment probabilities far from 0 or 1 enforcing the respect of the overlap condition which is necessary for valid inference.
- Number of covariates included : The more covariates included to personalize the allocation strategy, the greater the number of subgroups that need to be explored. While personalization has the potential to improve outcomes, too many heterogeneity dimensions will require more observations to be learnt.

Balancing exploration and exploitation and the selection of appropriate design hyperparameters is an ongoing challenge in adaptive experiments.

Choosing the best metric We choose to optimize the probability of visiting the profile webpage. However, a more exhaustive proxy for user engagement could have been picked. For example, tracking the frequency of profile modifications or the publication of profiles could have captured a broader range of user behaviors and indicated a higher level of engagement. By focusing solely on the probability of visiting the webpage, we may have overlooked the potential relationship between our interventions and these more comprehensive engagement proxies. It is possible that our interventions had an impact on user behaviors beyond the simple visit, and these behaviors could have been better captured by alternative measures. Choosing the appropriate engagement proxy involves a trade-off between the availability of data in a short term horizon and the comprehensiveness of the measure. The use of surrogates in this setting might be an interesting area of research.

Non-stationary relation between covariates and outcome Our adaptive experiment design, exploiting the responses of job seekers to treatments by batch over time, relied on contextual bandit methods which typically assume stationarity of the data. We decided not to model this temporal dimension, which may hide certain usage or seasonal trends. However, base job seeker engagement with the PES's online tools may vary across time (e.g. job seekers may less often visit the website on holidays), regardless of their treatment by an email allocation policy. This non-stationarity may in turn create a form of hidden variable bias, and fragilize contextual bandit strategies. The handling of the temporal dimension in adaptive experiments is a technical challenge in non-stationary environments. Practical methods for controlling its effect while maintaining favorable conditions for causal inference and evaluation are still lacking in our knowledge.

8 Conclusion

In this experiment, our objective was to increase the usage of an online platform developed by the French Public Employment Agency Pôle emploi. We did so by sending emails with varying content sequentially to jobseekers. We sought to understand the impact of these emails on users' engagement and assess the potential for personalization based on individual characteristics. We used an adaptive design in an attempt to increase take-up rates within the experiment, while gathering information on most promising emails.

Our results show that emails have a positive effect on the use of the platform, as evidenced by various metrics such as visits on the profile page, modification and publication of the profile. A deeper analysis into population subgroups shows that individuals who were more likely to respond to the incentives tended to be more experienced and qualified jobseekers who already had a closer connection to the institution.

The adaptive design did not allow to increase overall take-up rates within the experiment. Moreover, no email content clearly outperformed the others. Moreover, an optimal personalized allocation policy estimated on collected data could not outperform a random allocation.

These results indicate that the primary effect of the emails was to inform and remind jobseekers about the existence and purpose of the skills profile, regardless of the specific content. The success of the most minimalist email reinforces this interpretation.

References

ABEBE, G., A. S. CARIA, M. FAFCHAMPS, P. FALCO, S. FRANKLIN, AND S. QUINN (2020): "Anonymity or Distance? Job Search and Labour Market Exclusion in a Growing African City," *The Review of Economic Studies*, 88, 1279–1310.

ABEL, M., R. BURGER, AND P. PIRAINO (2020): "The value of reference letters: Experimental evidence from south africa," *American Economic Journal: Applied Economics*, 12, 40–71.

AGRAWAL, S. AND N. GOYAL (2013): "Thompson sampling for contextual bandits with linear payoffs," in *International conference on machine learning*, PMLR, 127–135.

ATHEY, S., U. BYAMBADALAI, V. HADAD, S. K. KRISHNAMURTHY, W. LEUNG, AND J. J. WILLIAMS (2022): "Contextual Bandits in a Survey Experiment on Charitable Giving: Within-Experiment Outcomes versus Policy Learning," *arXiv preprint arXiv:2211.12004*.

ATHEY, S., J. TIBSHIRANI, S. WAGER, AND OTHERS (2019): "Generalized random forests," *The Annals of Statistics*, 47, 1148–1178, publisher: Institute of Mathematical Statistics.

BAICKER, K., W. J. CONGDON, AND S. MULLAINATHAN (2012): "Health Insurance Coverage and Take-Up: Lessons from Behavioral Economics," *The Milbank Quarterly*, 90, 107–134.

BALLIET, D. AND P. A. VAN LANGE (2013): "Trust, conflict, and cooperation: a meta-analysis." *Psychological bulletin*, 139, 1090.

BANDURA, A. (1997): *Self-efficacy: The exercise of control*, W. H. Freeman.

BASSI, V. AND A. NANSAMBA (2018-01): "Screening and Signaling Non-Cognitive Skills: Experimental Evidence from Uganda," *SSRN Electronic Journal*.

BATTOCCHI, K., E. DILLON, M. HEI, G. LEWIS, P. OKA, M. OPRESCU, AND V. SYRGKANIS (2019): "EconML: A Python Package for ML-Based Heterogeneous Treatment Effects Estimation," https://github.com/py-why/EconML, version 0.x.

BETTINGER, E. P., B. T. LONG, P. OREOPOULOS, AND L. SANBONMATSU (2012): "The role of application assistance and information in college decisions: Results from the H&R Block FAFSA experiment," *The Quarterly Journal of Economics*, 127, 1205–1242, publisher: MIT Press.

BHARGAVA, S., G. LOEWENSTEIN, AND J. SYDNOR (2017): "Choose to Lose: Health Plan Choices from a Menu with Dominated Option," *The Quarterly Journal of Economics*, 132, 1319–1372, _eprint: https://academic.oup.com/qje/article-pdf/132/3/1319/30646793/qjx011.pdf.

BHARGAVA, S. AND D. MANOLI (2015): "Psychological Frictions and the Incomplete Take-Up of Social Benefits: Evidence from an IRS Field Experiment," *American Economic Review*, 105, 3489–3529.

BIBAUT, A., M. DIMAKOPOULOU, N. KALLUS, A. CHAMBAZ, AND M. VAN DER LAAN (2021): "Post-Contextual-Bandit Inference," in *Advances in Neural Information Processing Systems*, ed.

by M. Ranzato, A. Beygelzimer, Y. Dauphin, P. S. Liang, and J. W. Vaughan, Curran Associates, Inc., vol. 34, 28548–28559.

Bied, G., M. Hoffmann, C. Marie, and B. Picard (2023): "Signaling Skills Online and Labor Market Outcomes: Evidence from an Adaptive Experiment," AEA RCT Registry.

Bordalo, P., N. Gennaioli, and A. Shleifer (2022): "Salience," *Annual Review of Economics*, 14, 521–544.

Caria, S., M. Kasy, S. Quinn, S. Shami, A. Teytelboym, and others (2020): "An adaptive targeted field experiment: Job search assistance for refugees in Jordan," *Publisher: CESifo Working Paper.*

Carranza, E., R. Garlick, K. Orkin, and N. Rankin (2021): "Job Search and Hiring with Two-Sided Limited Information about Workseekers' Skills," *Institute of Labor Economics (IZA)*, type: IZA Discussion Papers.

Chernev, A., U. Böckenholt, and J. Goodman (2015): "Choice overload: A conceptual review and meta-analysis," *Journal of Consumer Psychology*, 25, 333–358.

Chernozhukov, V., M. Demirer, E. Duflo, and I. Fernández-Val (2019): "Generic Machine Learning Inference on Heterogenous Treatment Effects in Randomized Experiments," *arXiv.*

Chun, M. M., J. D. Golomb, and N. B. Turk-Browne (2011): "A Taxonomy of External and Internal Attention," *Annual Review of Psychology*, 62, 73–101.

Conway, A. R. A., N. Cowan, and M. F. Bunting (2001): "The cocktail party phenomenon revisited: The importance of working memory capacity," *Psychonomic Bulletin & Review*, 8, 331–335.

Dechêne, A., C. Stahl, J. Hansen, and M. Wänke (2010): "The Truth About the Truth: A Meta-Analytic Review of the Truth Effect," *Personality and Social Psychology Review*, 14, 238–257.

Deci, E. L. and R. M. Ryan (2000): "The "What" and "Why" of Goal Pursuits: Human Needs and the Self-Determination of Behavior," *Psychological Inquiry*, 11, 227–268.

Dhia, A. B., B. Crépon, E. Mbih, L. Paul-Delvaux, B. Picard, and V. Pons (2022): "Can a website bring unemployment down? Experimental evidence from France," Tech. rep., National Bureau of Economic Research.

Dhia, A. B. and E. Mbih (2020): "Do informational frictions affect enrollment in public-sponsored training? Results from an online experiment," *Job Market Paper.*

Diamond, A. (2013): "Executive Functions," *Annual Review of Psychology*, 64, 135–168.

Dimakopoulou, M., Z. Zhou, S. Athey, and G. Imbens (2018): "Estimation Considerations in Contextual Bandits," .

——— (2019): "Balanced linear contextual bandits," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, 3445–3453.

Dynarski, S., C. Libassi, K. Michelmore, and S. Owen (2021-06): "Closing the Gap: The Effect of Reducing Complexity and Uncertainty in College Pricing on the Choices of Low-Income Students," *American Economic Review*, 111, 1721–56.

Finkelstein, A. and M. J. Notowidigdo (2019): "Take-Up and Targeting: Experimental Evidence from SNAP," *The Quarterly Journal of Economics*, 134, 1505–1556.

Hadad, V., D. A. Hirshberg, R. Zhan, S. Wager, and S. Athey (2021): "Confidence intervals for policy evaluation in adaptive experiments," *Proceedings of the national academy of sciences*, 118.

Hallsworth, M. (2023): "A manifesto for applying behavioural science," *Nature Human Behaviour*, 7, 310–322.

Janssens, J. and N. Van Mechelen (2022): "To take or not to take? An overview of the factors contributing to the non-take-up of public provisions," *European Journal of Social Security*, 24, 95–116.

Jones, S. and K. Sen (2022): "Labour market effects of digital matching platforms: Experimental evidence from sub-Saharan Africa," *IZA Discussion Paper*.

Kuhn, P. J. (2014): "The internet as a labor market matchmaker," *IZA World of Labor*.

Laibson, D. (1997): "Golden Eggs and Hyperbolic Discounting," *The Quarterly Journal of Economics*, 112, 443–478.

Linos, E., A. Prohofsky, A. Ramesh, J. Rothstein, and M. Unrath (2022): "Can Nudges Increase Take-Up of the EITC? Evidence from Multiple Field Experiments," *American Economic Journal: Economic Policy*, 14, 432–52.

Mani, A., S. Mullainathan, E. Shafir, and J. Zhao (2013): "Poverty Impedes Cognitive Function," *Science*, 341, 976–980.

Marinescu, I. and R. Wolthoff (2020): "Opening the black box of the matching function: The power of words," *Journal of Labor Economics*, 38, 535–568.

Michie, S., M. M. van Stralen, and R. West (2011): "The behaviour change wheel: A new method for characterising and designing behaviour change interventions," *Implementation Science*, 6, 42.

Moscovici, S. (2008): *Psychoanalysis: Its image and its public*, Polity Press.

Offer-Westort, M., A. Coppock, and D. P. Green (2021): "Adaptive experimental design: Prospects and applications in political science," *American Journal of Political Science*, 65, 826–844.

Pallais, A. (2014): "Inefficient Hiring in Entry-Level Labor Markets," *American Economic Review*, 104, 3565–99.

Rubin, D. B. (1974): "Estimating causal effects of treatments in randomized and nonrandomized studies." *Journal of educational Psychology*, 66, 688.

Ryan, R. M. and E. L. Deci (2000): "Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being." *American Psychologist*, 55, 68–78.

Schanzenbach, D. W. (2009): "Experimental Estimates of the Barriers to Food Stamp Enrollment," *Institute for Research on Poverty*.

Scheibehenne, B., R. Greifeneder, and P. M. Todd (2010-10): "Can There Ever Be Too Many Options? A Meta-Analytic Review of Choice Overload," *Journal of Consumer Research*, 37, 409–425.

Su, Y., M. Dimakopoulou, A. Krishnamurthy, and M. Dudík (2020): "Doubly robust off-policy evaluation with shrinkage," .

Sutton, R. S. and A. G. Barto (2018): *Reinforcement learning: An introduction*, MIT press.

Sverdrup, E., A. Kanodia, Z. Zhou, S. Athey, and S. Wager (2020): "policytree: Policy learning via doubly robust empirical welfare maximization over trees," *Journal of Open Source Software*, 5, 2232.

Thompson, W. R. (1933): "On the likelihood that one unknown probability exceeds another in view of the evidence of two samples," *Biometrika*, 25.

Tversky, A. and D. Kahneman (1973-09): "Availability: A heuristic for judging frequency and probability," *Cognitive Psychology*, 5, 207–232.

Villar, S. S., J. Bowden, and J. Wason (2015): "Multi-armed bandit models for the optimal design of clinical trials: benefits and challenges," *Statistical science: a review journal of the Institute of Mathematical Statistics*, 30, 199.

Wager, S. (2022): "STATS 361: Causal Inference," Compiled notes from the Spring 2022 version of the class, Chapter "Lecture 14: Adaptive Experiments," accessed on September 24, 2023.

Wagner, W., G. Duveen, R. Farr, S. Jovchelovitch, F. Lorenzi-Cioldi, I. Marková, and D. Rose (1999): "Theory and Method of Social Representations," *Asian Journal of Social Psychology*, 2, 95–125.

Wheeler, L., R. Garlick, E. Johnson, P. Shaw, and M. Gargano (2022): "LinkedIn(to) Job Opportunities: Experimental Evidence from Job Readiness Training," *American Economic Journal: Applied Economics*, 14, 101–25.

Woolley, K. and A. Fishbach (2017): "Immediate Rewards Predict Adherence to Long-Term Goals," *Personality and Social Psychology Bulletin*, 43, 151–162.

——— (2018): "It's about time: Earlier rewards increase intrinsic motivation." *Journal of Personality and Social Psychology*, 114, 877–890.

Zhan, R., V. Hadad, D. A. Hirshberg, and S. Athey (2021): "Off-Policy Evaluation via Adaptive Weighting with Data from Contextual Bandits," .

A Individual characteristics included in the algorithm

We collect several information on jobseekers pre-treatment. These data are used to constitute a matrix X of individual characteristics to target the treatments and personalize the allocation and are presented in table 7.

Socio-demographic variables	
Age	numeric
Male	binary
Type of employment area[14]	categorical (x10)
Number of children	numeric
Familial situation	categorical (x6)
Highest diploma's level	categorical (x4)
Targeted job	
Reservation wage	numeric
Years spent working in the sector of interest	numeric
Number of different target jobs declared on the online profile	numeric
The jobseeker is looking for a full-time job	categorical (x2)
Level of qualification of the individual for the targeted job	categorical (x4)
Target job sector	categorical (x14)
Target type of contract	categorical (x4)
Characterization of experienced unemployment spells	
Number of days since last registration on Pôle emploi's lists	numeric
Total duration of unemployment in lifetime in days registered by Pôle emploi	numeric
Number of unemployment periods in lifetime registered by Pôle emploi	numeric
Reason why the jobseeker registered at Pôle emploi	categorical (x15)
Pôle emploi's administrative category[15]	categorical (x6)
Number of days before the person is no longer eligible to unemployment benefits[16]	numeric
Wants to become an entrepreneur	categorical(x2)
The jobseeker declared having a non advanced level of French	binary
Relationships with the agency	
Number of interviews with the caseworker during the last 3 months	numeric
Type of accompaniment received from Pôle emploi	categorical (x4)
Main obstacles assumed to slow return to employment	categorical (x4)
Proxy for the distance between home and agency	numeric
Receives any benefits (whether unemployed or not)	categorical (x3)
Usage of Pôle emploi's digital services	
Already published its online profile	binary
Number of days during which the individual connected on his/her personal account during the last month	numeric
Others	
Number of distinct recruiters who connected on their account and looked for profiles on the platform during the last month in the employment pool[17]	numeric

Table 7: Individual characteristics

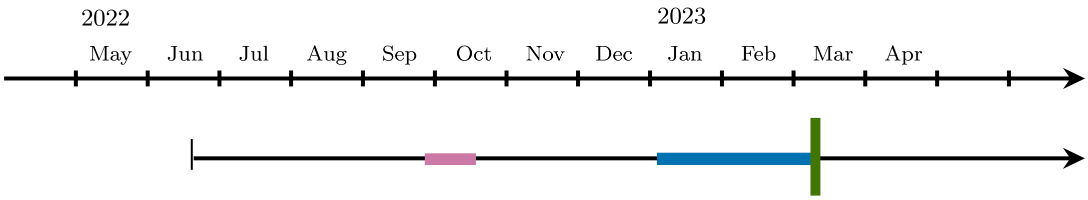
[15]the categories of jobseekers take into account the situation of the unemployed person (unemployed, in training, on sick leave, etc.) and his or her objectives (fixed-term contract, open-ended contract, full time, etc.)

[16]negative if the benefits stopped before the sampling day

[17]An employment pool is an administrative geographic area within which most of the active population resides and works, and in which firms can find most of the labor needed to fill the job positions. This division is adapted to local labor market studies.

B Experiment timeline

Figure 11: Timeline of the experiment



Pilot
Objectives: (1) Test the algorithm on real data, verify the automation of the assignment to treatment groups and the sending of emails (2) Check the relevance of the outcome

Experiment
252 000 individuals, among which 84 000 are treated
Objectives: (1) Sending the best emails to people depending on their individual characteristics (2) Evaluate the impact of each block and of each email

Last Batch: Optimal policy versus random allocation
76 000 individuals, half receive an email based on the policy tree, and the others receive one email at random
Objectives: Find the optimal personalized policy and determine whether or not this estimated policy has more impact than an random allocation

C Emails components

C.1 Information on Recruiters and on *Pôle emploi*'s actions

C.1.1 I1 - Information on recruiters

Make yourself visible to recruiters

350,000 recruiters search for candidates on our platform every month. Your profile may interest them, publish it so they can contact you easily!

C.1.2 I2 - Information on *Pôle emploi*'s actions

Help us to better support you

Filling in your profile allows you to take stock of your skills and helps us supporting you in your professional project. Our automatic job and training recommendations become more relevant and better targeted.

C.2 H1 - Tutorial

How to fill in your profile?

1. Log in to your personal space and click on "My skills profile and my CV"
2. In "Career path" add your experiences and trainings
3. In "Skills" indicate your knowledge, what you can do and your personal qualities
4. In "Resume and Achievements" add your CV and all the useful documents to share in order to showcase yourself
5. In "Jobs and Projects", indicate your professional project
6. Click on the "make visible" button if you want your profile to be visible to recruiters

C.3 H2 - Intensive Help

C.3.1 H2w - Workshop

A workshop to support you

If you wish to be accompanied in the process, you can register for the half-day workshop "Reviewing my skills". This workshop will allow you to:

- Identify your skills
- Learn how to promote them on the job market
- Complete your profile and make it visible to recruiters

To register, click on the button below and go to the “Prepare your application” section. You will find a link to the workshop.

Register for the workshop

C.3.2 H2c - Caseworker

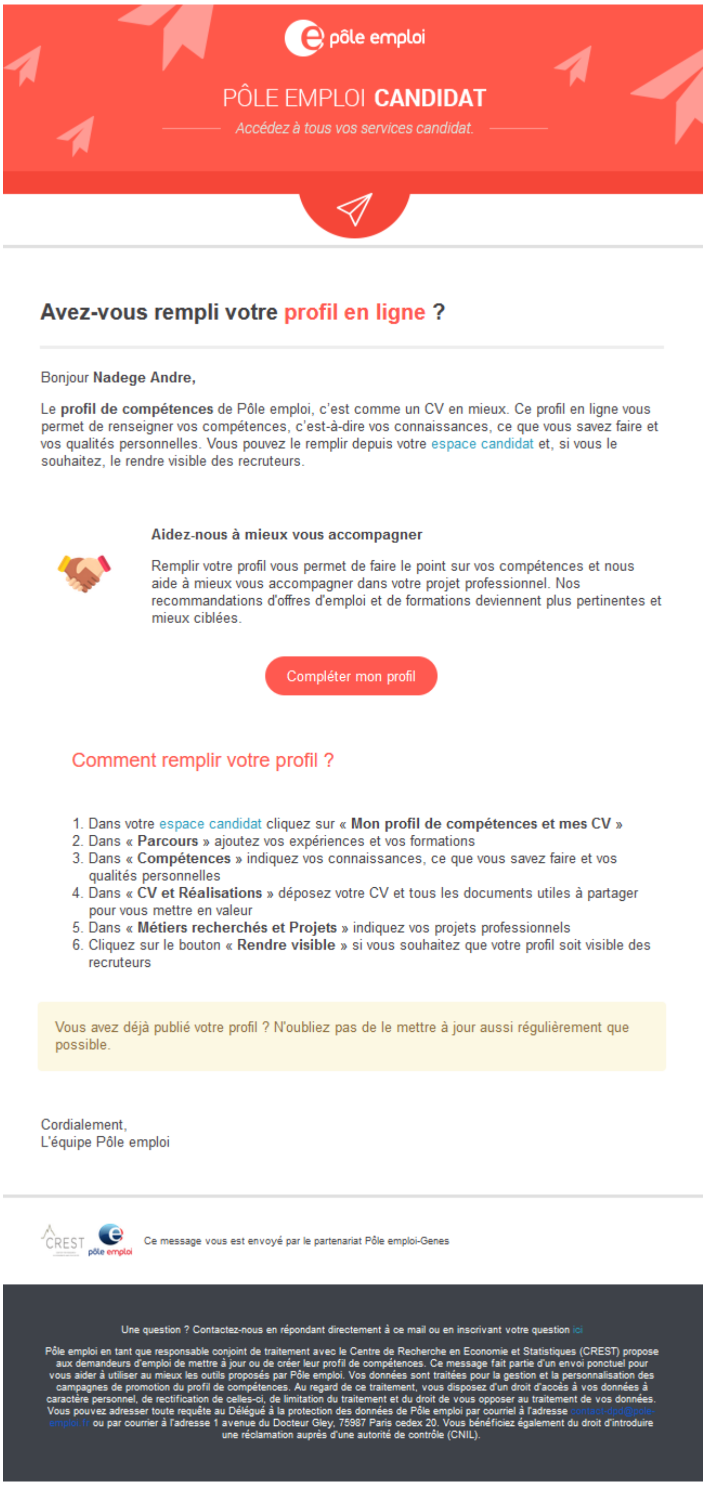
Talk about it with your caseworker

If you wish to be accompanied in the process, you can book an appointment by clicking below:

Book an appointment

C.4 Email example

Figure 12: Example of one email



pôle emploi

PÔLE EMPLOI **CANDIDAT**

Accédez à tous vos services candidat.

Avez-vous rempli votre profil en ligne ?

Bonjour **Nadege Andre,**

Le **profil de compétences** de Pôle emploi, c'est comme un CV en mieux. Ce profil en ligne vous permet de renseigner vos compétences, c'est-à-dire vos connaissances, ce que vous savez faire et vos qualités personnelles. Vous pouvez le remplir depuis votre espace candidat et, si vous le souhaitez, le rendre visible des recruteurs.

Aidez-nous à mieux vous accompagner

Remplir votre profil vous permet de faire le point sur vos compétences et nous aide à mieux vous accompagner dans votre projet professionnel. Nos recommandations d'offres d'emploi et de formations deviennent plus pertinentes et mieux ciblées.

Compléter mon profil

Comment remplir votre profil ?

1. Dans votre espace candidat cliquez sur « **Mon profil de compétences et mes CV** »
2. Dans « **Parcours** » ajoutez vos expériences et vos formations
3. Dans « **Compétences** » indiquez vos connaissances, ce que vous savez faire et vos qualités personnelles
4. Dans « **CV et Réalisations** » déposez votre CV et tous les documents utiles à partager pour vous mettre en valeur
5. Dans « **Métiers recherchés et Projets** » indiquez vos projets professionnels
6. Cliquez sur le bouton « **Rendre visible** » si vous souhaitez que votre profil soit visible des recruteurs

Vous avez déjà publié votre profil ? N'oubliez pas de le mettre à jour aussi régulièrement que possible.

Cordialement,
L'équipe Pôle emploi

CREST pôle emploi Ce message vous est envoyé par le partenariat Pôle emploi-Genes

Une question ? Contactez-nous en répondant directement à ce mail ou en inscrivant votre question ici

Pôle emploi en tant que responsable conjoint de traitement avec le Centre de Recherche en Economie et Statistiques (CREST) propose aux demandeurs d'emploi de mettre à jour ou de créer leur profil de compétences. Ce message fait partie d'un envoi ponctuel pour vous aider à utiliser au mieux les outils proposés par Pôle emploi. Vos données sont traitées pour la gestion et la personnalisation des campagnes de promotion du profil de compétences. Au regard de ce traitement, vous disposez d'un droit d'accès à vos données à caractère personnel, de rectification de celles-ci, de limitation du traitement et du droit de vous opposer au traitement de vos données. Vous pouvez adresser toute requête au Délégué à la protection des données de Pôle emploi par courriel à l'adresse contact-dpd@pole-emploi.fr ou par courrier à l'adresse 1 avenue du Docteur Gley, 75987 Paris cedex 20. Vous bénéficiez également du droit d'introduire une réclamation auprès d'une autorité de contrôle (CNIL).

D Balance checks

Table 8: Sample characteristics

	Control mean	Treated mean	p-value
Male	0.490	0.487	0.111
Age	39.830	39.868	0.541
Number of days since the inscription	574.051	576.546	0.602
Highest diploma's level			
Others	0.153	0.153	0.222
University degree	0.326	0.326	0.865
End-of-high-school diploma	0.231	0.231	0.607
Vocational degree	0.290	0.290	0.976
Level of qualification			
Skilled worker	0.144	0.143	0.555
Unqualified employee	0.212	0.211	0.453
Qualified employee	0.494	0.495	0.738
Executive	0.150	0.151	0.418
Type of contract			
Others	0.085	0.083	0.011**
Temporary contract	0.232	0.235	0.000***
Permanent contract	0.684	0.682	0.026**
Type of accompaniment[1] received from Pôle emploi			
Others	0.035	0.036	0.793
Reinforced	0.186	0.186	0.977
Guided	0.542	0.543	0.734
Follow-up	0.236	0.235	0.561

Notes: Table 1 shows summary statistics for the sample of 252, 000 jobseekers. Column (1) and (2) present the mean in the control and the treatment group respectively. Column (3) presents p-values based on regressions that include batch fixed effects and heteroskedasticity-robust standard errors clustered by batch.

[1] (1) Reinforced: Intended for people who require intensive support. This modality is based on more frequent contacts with caseworkers, and face-to-face meetings are preferred. The number of jobseekers who can be accompanied by a caseworker is limited to 70. (2) Guided: Dedicated to jobseekers in an intermediate situation. The number of jobseekers who can be accompanied by a caseworker is 100 to 150. (3) Follow-up: Intended for jobseekers who are closest to the job market and have the greatest autonomy. Dematerialized contact methods (phone and email) are preferred for exchanges with their caseworkers. The number of jobseekers who can be accompanied by a caseworker is 200 to 300.

E Complementary analysis

E.1 Impacts under linearity assumption

In a complementary analysis, we consider the 38,000 individuals treated by emails variants chosen uniformly at random during the run-off, and estimate regression:

$$Y_i^{obs} = \beta_0 + \sum_k \beta_k 1\{l_{ik} = 1\} + \epsilon_i \tag{7}$$

where l_{ik} is a binary indicator denoting that individual i received an email in which communication variant k was present.

Results of this regression are presented in Table 9. All coefficients β_k for the effect of the presence of the communication variants are slightly negative; none are significant at the 0.01 threshold. In other words, in this linear specification, no communication variant improves on the shortest and simplest e-mail. A F-test for the joint equality and nullity of the β_k's does not allow the rejection of the null hypothesis ($p = 0.146$).

Table 9: Impact of behavioral lever inclusion under linearity assumption

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	0.0501***	0.003	18.948	0.000	0.045	0.055
Information 1	-0.0003	0.002	-0.122	0.903	-0.004	0.004
Information 2	-0.0027	0.002	-1.243	0.214	-0.007	0.002
Tutorial	-0.0007	0.003	-0.241	0.809	-0.007	0.005
Workshop	-0.0003	0.003	-0.103	0.918	-0.006	0.006
Caseworker	-0.0062**	0.003	-2.034	0.042	-0.012	-0.000

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

This table shows the results from equation (7) that assumes additive effects of the different components of the emails. The reported coefficients are estimated on 38 000 individuals who randomly received one of the 16 emails. The estimation is conducted on 252 000 individuals from the main experiment.

F Robustness check - Generic Machine Learning

F.1 Classification Analyses - Other variables

	20% Least	20% Most	Difference	pvalue
Experience	2.74	11.08	8.33	0.00
Reservation wage	1933.42	2299.65	359.69	0.00
Category				
jobseeker (Required to Search for Work)	0.63	0.65	0.02	0.00
Partially Employed jobseeker	0.06	0.12	0.06	0.00
Part-time jobseeker	0.18	0.16	-0.02	0.00
Unavailable for Work	0.03	0.03	0.00	0.07
Employed (Not Required to Search for Work)	0.11	0.04	-0.08	0.00
Type of accompaniment[1] received from Pôle emploi				
Follow-up	0.28	0.24	-0.04	0.00
Guided	0.58	0.53	-0.05	0.00
Reinforced	0.11	0.18	0.06	0.00
Others	0.05	0.04	-0.00	0.00
Training level				
End-of-high-school diploma	0.28	0.20	-0.08	0.00
Others	0.10	0.16	0.06	0.00
Vocational degree	0.26	0.31	0.06	0.00
University degree	0.37	0.34	-0.03	0.00
Sector				
Agriculture	0.05	0.03	-0.02	0.00
Arts, Artistic Crafts	0.01	0.01	-0.00	1.00
Banking, Insurance, Real Estate	0.02	0.02	-0.00	0.27
Commerce, Sales, and Retail	0.15	0.13	-0.03	0.00
Communication, Media, and Multimedia	0.03	0.02	-0.01	0.00
Construction, Building, and Public Works	0.07	0.06	-0.01	0.00
Restaurant Industry, Tourism, Leisure	0.08	0.07	-0.01	0.00
Industry	0.07	0.07	-0.00	0.06
Installation and Maintenance	0.04	0.04	-0.00	0.59
Health	0.04	0.03	-0.01	0.00
Personal and Community Services	0.15	0.20	0.06	0.00
Entertainment	0.04	0.03	-0.01	0.00
Business Support	0.12	0.20	0.07	0.00
Transportation and Logistics	0.11	0.09	-0.02	0.00
Other	0.02	0.01	-0.01	0.00
Contract type				
Permanent contract	0.63	0.73	0.10	0.00
Temporary contract	0.26	0.20	-0.06	0.00
Others	0.03	0.04	0.02	0.00
Main obstacle diagnosed to return to employment				
Professional project	0.20	0.19	-0.01	0.00
Direct return	0.17	0.16	-0.01	0.05
Freins	0.08	0.13	0.05	0.00
Research	0.34	0.38	0.04	0.00

Table 10: Classification analysis

Note: The following table presents the average characteristics of the groups classified as the least and most affected. The table also includes the results of systematic tests conducted to examine the differences between these two groups.

[1] (1) Reinforced: Intended for people who require intensive support. This modality is based on more frequent contacts with caseworkers, and face-to-face meetings are preferred. The number of jobseekers who can be accompanied by a caseworker is limited to 70. (2) Guided: Dedicated to jobseekers in an intermediate situation. The number of jobseekers who can be accompanied by a caseworker is 100 to 150. (3) Follow-up: Intended for jobseekers who are closest to the job market and have the greatest autonomy. Dematerialized contact methods (phone and email) are preferred for exchanges with their caseworkers. The number of jobseekers who can be accompanied by a caseworker is 200 to 300.

	β_{ATE}	β_{HET}
Estimate	0.0245	1.08
Conf. interv.	[0.0217, 0.0272]	[0.709, 1.44]
Corrected pvalue	0.000	0.000

Table 12: ATE and HTE estimates (Elastic Net)

	Least affected (G_1)	Most affected (G_5)	Difference (G_5- G_1)
Estimate	0.0118	0.0517	0.0394
Conf. interv.	[0.00643 , 0.0171]	[0.0430, 0.0593]	[0.0306, 0.0489]
Corrected pvalue	0.007	0.000	0.000

Table 13: GATES estimates for least and most affected groups (Elastic Net)

F.2 Comparison of Machine Learning Proxy

Table 11 compares four ML methods for producing proxy predictors using criterion proposed in Chernozhukov et al. (2019). Generalized causal forest outperform all methods and is used in the core of the paper. The second best ML predictor is Elastic Net.

	Elastic Net	Gradient Boosting	Generalized Random Forest	Random Forest
Best BLP Λ	0.017593	0.015643	**0.018799**	0.011611
Best GATES $\bar{\Lambda}$	0.000787	0.000796	**0.000855**	0.00070

Table 11: Comparison of the quality of the ML predictors

F.3 Results for the second best predictor

In this section we provide the same analysis given in section 4.3 but using the second best machine learning predictor, Elastic Net. Results are consistent and do not alter our analysis.

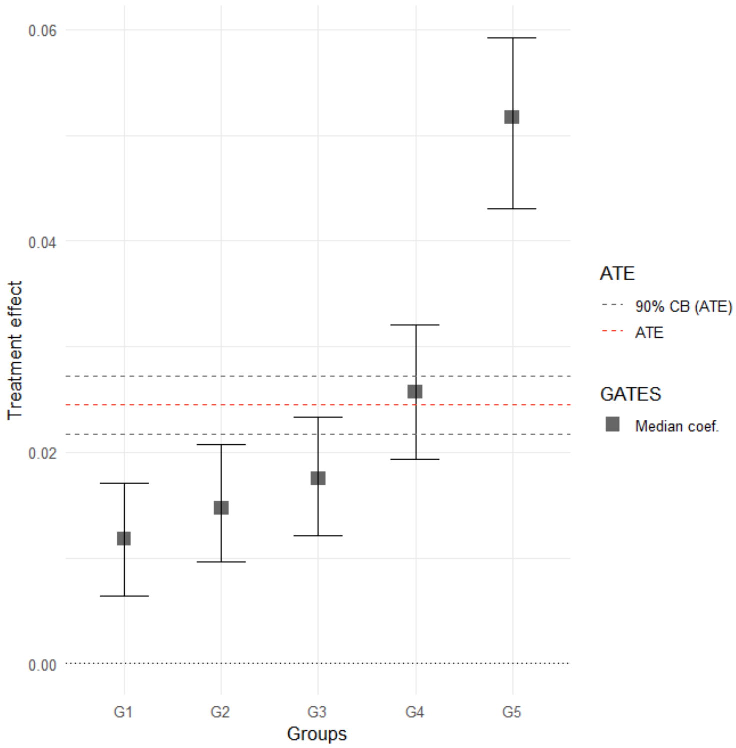


Figure 13: GATES (Elastic Net)

F.4 Classification Analyses - Main variables - Elastic Net

	20% Least	20% Most	Difference	pvalue
Experience				
Below 5 years	0.87	0.31	-0.56	0.00
Between 5 and 10 years	0.09	0.13	0.04	0.00
Above 10 years	0.04	0.56	0.53	0.00
Reservation Wage				
Below 25	0.29	0.18	-0.11	0.00
Between 25 and 50	0.24	0.15	-0.09	0.00
Between 50 and 75	0.23	0.25	0.03	0.00
Above 75	0.24	0.41	0.16	0.00
Qualification Level				
Skilled Worker	0.19	0.11	-0.08	0.00
Non qualified employee	0.23	0.16	-0.07	0.00
Qualified employee	0.48	0.52	0.04	0.00
Executive	0.11	0.21	0.11	0.00
Age				
Below 25	0.31	0.01	-0.31	0.00
Between 25 and 39	0.52	0.12	-0.39	0.00
Between 40 and 54	0.11	0.38	0.27	0.00
Above 55	0.01	0.49	0.49	0.00
Gender				
Is male	0.54	0.45	-0.09	0.00
Relationship with institution				
Already published a profile	0.19	0.22	0.02	0.00
At least one visit in the last month	0.11	0.62	0.51	0.00
At least one meeting in the last month	0.27	0.57	0.30	0.00
Number of days since last registration				
Below 3 months	0.27	0.20	-0.08	0.00
Between 3 months and 1 year	0.73	0.80	0.08	0.00
Between 1 year and 2 years	0.31	0.48	0.17	0.00
Above 2 years	0.13	0.30	0.17	0.00

Note: The following table presents the average characteristics of the groups classified as the least and most affected. The table also includes the results of systematic tests conducted to examine the differences between these two groups.

Table 14: Classification analysis (Elastic Net)

G Algorithm

G.1 Detailed procedure

We follow Dimakopoulou et al. 2019 and use a Generalized Random Forest Thompson sampling algorithm in order to assign individuals to treatment groups at each batch.

We have a finite set of available treatments $\mathcal{W} = \{1, \ldots, W\}$ which we want to assign optimally to each jobseeker. For each of them we observe a context vector $x_i \in \mathbb{R}^d$, the treatment assignment $w_i \in \mathcal{W}$ and observe the response to this treatment $y_i^{\text{obs}} \in \{0, 1\}$. jobseekers in the treated group are observed by batch, we denote $\mathcal{B} = \{1, ..., B\}$ the set of equally sized batches.

Let s_β be the set of individuals in batch number β. $\{(Y_i^{\text{obs}}, W_i, X_i)\}_{i \in s_\beta}$ are thus the outcomes, indicators for the treatment received and the characteristics of each individual in this batch. After running the intervention for $b \in \mathcal{B}$ batches, collected data constitute the history which is a set of random variables $H^b = \{(Y_i^{\text{obs}}, W_i, X_i)\}_{\{i \in s_\beta, \ \forall \beta \in [\![1,b]\!]\}}$ taking real values $h^b = \{\{(y_i^{\text{obs}}, w_i, x_i)\}_{i \in s_\beta}\}_{\beta \in [\![1,b]\!]}$. Thus, the history contains the information of every individuals in $S_b = \bigcup_{\beta=1}^{b} s_b$.

In the first batch, $b = 1$, we assign treatments randomly with a probability $\frac{1}{W}$. This first batch allow us to collect the first set of covariates and response pairs.

For all the following batches $b = 2, .., B$:

> Note : It is important to notice that unlike Dimakopoulou et al. (2019), where several models are trained on responses for each treatment arm, there is only one model where treatments are included in the covariates as dummy variables.

1. For each observation in batch b, namely for $i \in s_b$

   (a) Predict $\hat{\mu}_w(x_i)$ and $V(\hat{\mu}_w(x_i))$ for each treatment arm w in $\mathcal{W}$ using the model.

   (b) For each arm w, draw M samples from the following normal distribution :

   $$\hat{Y}_{i,w}^m \sim \mathcal{N}\left(\hat{\mu}_w(x_i), V(\hat{\mu}_w(x_i))\right) \text{ for } m = 1, ..., M$$

   (c) Compute :

   $$p_{i,w} = \frac{1}{M} \sum_{m=1}^{M} \mathbb{1}\left\{w = \arg\max_{w' \in \mathcal{W}} \hat{Y}_{i,w}^m\right\} \qquad \forall w \in \mathcal{W}$$

   where $p_{i,w}$ are frequencies computed by counting how many times treatment w produces the highest estimate of the success probability over all counterfactual treatments on M drawings.

   (d) We define a probability floor $e_{min} > 0$ so that the probability of being assigned to any treatment is bounded away from zero or one. This minimum probability ensures that we keep exploring all the treatments during the experiment, even if it appeared to perform poorly in the first periods. It is also preferable to keep non-zero sampling probabilities for the inference. Thus, we replace any $p_{i,w} < e_{min}$ by e_{min}. Since we need all probabilities to sum to 1, we rescale to get the final probabilities e_w^b for batch b with the following rule:

$$e_{i,w} = \frac{p_{i,w}}{\sum_{w \in \mathcal{W}} p_{i,w}}$$

(e) We assign treatments according to the probabilities, which amounts to draw from a multinomial distribution with 1 trial and the probabilities $(e_w)_{w \in \mathcal{W}}$, in other words:

$$W_i^b \sim \text{Multinom}_1(e_{i,1}, ..., e_{i,W})$$

2. We collect the response y_i^{obs} and save it in the historical data of batch b: $\{(y_i, w_i, x_i)\}_{i \in s_b}$.

G.2 Hyper parameters and technical details

We set the minimum threshold for exploration $e_{min} = 0.005$, which remained fixed throughout the 7 rounds. Moreover, the number of draws from the normal distribution was set to $M = 1000$.

The GRF model has been implemented using the python package econml Battocchi et al. (2019) and the function

```
RegressionForest
```

Table 15 presents the hyperparameters that were tuned during the training of the GRF model, including their corresponding grid values.

Hyperparameter	**Grid**
max_features	[# Features / 2, $\sqrt{\#\text{Features}}$, # Features]
min_samples_leaf	[5, 10, 50, 100]
min_balancedness_tol	[0.4, 0.45, 0.485]
n_estimators	[2000]

Table 15: Algorithm hyper parameters

H Policy tree

Algorithm 1 Estimate potential rewards using the AIPW estimator

---

Require: Batches B_t of the data for all t, including X_t, W_t and Y_t
Ensure: Concatenated matrix $\hat{G}$ of all $\hat{G}_{B_t}$

Initialize empty matrix $\hat{G}$
for each batch B_t **do**
    Train a causal forest on batches $B_1, B_2, ...B_{t-1}$, predicting causal effects on $Y_1, ...Y_{t-1}$ of the different treatments depending on $X_1, ...X_{t-1}$
    Predict all $\hat{\mu}_i(x, w)$ on batch B_t based on this causal forest - taking the predictions instead of the causal effects
    for each component (i, w) of $\hat{G}_{B_t}$ **do**
        Estimate $\hat{G}_i(w)$ using the Augmented Inverse Probability Weighting estimator as follows:

$$\hat{G}_i(w) = \hat{\mu}_i(x, w) + \frac{\mathbb{1}(W_i = w)}{\tilde{e}_i\left(X_i, w\right)}\left(Y_i - \hat{\mu}_i(x, w)\right)$$

Note: $\tilde{e}_i\left(X_i, w\right)$ are fully known since they are determined by the design
    end for
    Append $\hat{G}_{B_t}$ to $\hat{G}$
end for

Algorithm 2 Policy Tree

Require: Concatenated matrix $\hat{G}$, covariates X
Ensure: Policy Tree T

Clip the components of $\hat{G}$ to remove outliers
Split the dataset into a train and a test set
for each depth d of the tree **do**
    Train a policy tree T with depth d on the train set, using the concatenated matrix $\hat{G}$ and covariates X
    Evaluate the performance of T on the test set
end for
Select the depth with the best performance
Train the policy tree T on the full dataset using the selected depth, using the concatenated matrix $\hat{G}$ and covariates X
