How can AI improve search and matching? Evidence from 59 million personalized job recommendations

Thomas Le Barbanchon
Lena Hensvik
Roland Rathelot *

July 2023, *Preliminary version*

Abstract

We explore how Artificial Intelligence can be leveraged to help frictional markets to clear. We design a collaborative-filtering machine-learning job recommender system that uses job seekers' click history to generate relevant personalised job recommendations. We deploy it at scale on the largest online job board in Sweden, and design a clustered two-sided randomised experiment to evaluate its impact on job search and labor-market outcomes. Combining platform data with unemployment and employment registers, we find that treated job seekers are more likely to click and apply to recommended jobs, and have 0.6% higher employment within the 6 months following first exposure to recommendations. At the job-worker pair level, we document that recommending a vacancy to a job seeker increases the probability to work at this workplace by 5%. Leveraging the two-sided vacancy-worker randomisation or the market-level randomisation, we find limited congestion effects. We find that employment effects are larger for workers that are less-educated, unemployed, and have initially a large geographic scope of search, for jobs that are attached to several jobs, and are relatively older. Results also suggest that recommendations expanding the occupational scope yield higher effects.

*Le Barbanchon: Bocconi University. Hensvik: Uppsala University. Rathelot: Institut Polytechnique de Paris. We would like to thank participants to ASSA, IZA, TEPP, ETH-UNIL-UZH Job search workshops, and to seminars in Essex, Tilburg, ASSA 2022, Tinbergen Institute, Maastricht, VfS Basel 2022, HEC Paris, Berlin, Princeton, Bonn, USC, UCLA Anderson, Stockholm Econ, UCL, Oxford, Copenhagen, Chicago Booth, CEPR Labor symposium 2023, and members of the IRS at Princeton for helpful comments. We thank Arbetsformedlingen for entering in this research partnership, and IFAU for providing access to computing facilities. Finally, we thank Ana Sofia Teles, Sara Rabino, Jeremy Marquis, Qian Dong, Guido Deiana, Agathe Rosenzweig, Vincenzo Alfano for superb research assistance. The financial support of the ERC StG 758190 "ESEARCH" is gratefully acknowledged. This study is registered in the AEA RCT Registry and the unique identifying number is: "AEARCTR-0003616". All errors are our own. Hensvik is also affiliated to IFAU, CEPR, and IZA, Le Barbanchon to CEPR, IGIER, IZA, J-PAL, Rathelot to CREST, CEPR, IZA, J-PAL.

1 Introduction

Over the last decade, there has been strong interest in the potential disruptive effect of Artificial Intelligence (AI) technology on various markets (Agrawal et al., 2019). The labor market is no exception.[1] As search and matching occur more and more frequently online, labor market intermediaries collect impressive amounts of data that can be used as inputs in AI models to develop tailored services. The rationale behind this approach is that online job platforms observe information about workers' and firms' search behaviors (and revealed preferences) that would help them to clear the market (as a central planner would do), lowering search costs and reducing mismatch (Milgrom and Tadelis, 2018). Online job platforms using advanced market-clearing technologies may deliver on the promise of an internet that solves information imperfections. However, little is known about the actual AI effects on labor market-clearing and whether current AI technologies are mature enough (Kircher, 2022).

In this paper, we provide the first comprehensive empirical analysis of the labor market effects of machine-learning job recommender systems. In partnership with *Arbetsförmedlingen*, the Swedish Public Employment Service (PES), we develop a recommender system that provides a personalized list of vacancies to every job seeker visiting the largest online job board in Sweden. The recommender system uses as input the naturally-occurring online data from the website activity. We use a clustered two-sided randomized controlled trial to evaluate the effects of recommendations on workers' search activity and matching outcomes. The scale of our experiment and the precision of our recommender system deliver new insights on the role of information imperfections on the labor market.

We leverage the data opportunities raised by the online job board *Platsbanken.se* maintained by the Swedish PES. Platsbanken comprises almost all vacancies posted in the Swedish labor market. We record job search activity, i.e., clicks/views of job ads and applications at the job seeker-job posting pair level. For evaluation purposes, we link the online search activity data of registered workers to employment and unemployment registers at the individual level. This allows us to estimate treatment effects on core labor market outcomes with a higher level of accuracy/precision than in studies analyzing interventions on private job boards.

First, we design a job recommender system. The recommender system takes the observed clicks/views data recorded on the website (bipartite graph between individual job seekers

[1]There is a recent and fast growing literature analysing how AI changes the type of tasks/jobs demanded in the labor market (see for example Acemoglu et al. (2022)). We provide a complementary approach focusing on how AI technology affects matching on the labor market.

and vacancies) as input, and delivers for each individual job seeker a list of recommendations ranked by relevance. The choice of the input data comes from legal and operational constraints that are common across many online job boards, and ensures the portability of our recommender system across websites and the external validity of our evaluation results. The algorithm leverages the fact that job seekers click on job ads they find interesting, and by doing so implicitly rate them. Job seekers who clicked on the same ads in the past have common preferences over jobs (i.e., unobserved latent factors). Broadly speaking, the system recommends to a given job seeker the job ads that job seekers with similar preferences viewed. From an economic point of view, the algorithm learns from the private experience of individual job seekers, and diffuses this information to other market participants. Hopefully, the recommender system leads to a reduction in information imperfections and generates mostly positive externalities across users. However, this remains an open question as the recommender system is essentially driven by a statistical objective without any explicit economic foundations.

Second, we design and implement a randomized controlled trial to estimate the effects of recommendations on matching. We randomize both job-seekers and vacancies (resp. 1.9 million users and 605,000 vacancies). We show treated job-seekers a personalized list of vacancy recommendations when they browse the *Platsbanken* website, while control job seekers are shown the default website, with no recommendations. Treated vacancies are included in recommendation lists, while control vacancies are not. Such an experimental design allows us to identify treatment effects on online search activity (clicks and applications) and on employment outcomes, but also on hiring rates from recruiters' perspective. The fact that we randomize both sides of the market also brings new identification power for congestion/displacement effect. We further complement the worker- and vacancy-level randomization with a *standard* market-level randomization: we partition the labor market into commuting zone $\times$ skill groups submarkets and leave a randomized subset of submarkets out of the experiment, as super control markets, which also allows us to detect displacement effects.

Treated job seekers increase by 44% the number of daily clicks for recommended vacancies, while they substitute away from non-recommended vacancies (-1%). Those opposing forces cancel out so that total clicks do not increase significantly. We find similar treatment effects on application behavior, suggesting that job seekers find the recommended vacancies relevant enough. However, the opposing forces result into a slight decrease in overall applications by 0.9% (statistically significant at the 10% level). Treated job seekers have higher employment rate by 0.65%, statistically significant at the 5% level.

From the recruiters' perspective, treated vacancies receive 1% more clicks and 2% more applications. The marginal clicks and applications come from treated users for whom the treated vacancies appeared in their recommendation set. We do not find evidence of large displacement effects from recruiters' perspective in the market-level randomization. We cannot detect differences in the number of clicks received by control vacancies in markets where some vacancies are treated vs. in super control markets, and the employment level of firms with control vacancies in treated market does not differ from that of firms in super control markets.

In a last step, we move towards a granular analysis of recommendation effects at the worker-recommended job pair level, using the 59 million recommendations in our sample. We show that our two-sided randomization plan allows to identify congestion/displacement effects in an innovative way. As the application behaviors of control users is the same towards control and treated jobs, any difference in their employment across jobs would be due to congestion effects. As treated workers apply more to treated jobs, control applicants to the same job face greater competition. Indeed, we find that employment of control workers is 1.4% lower in treated jobs. However, with standard errors of 1.7%, the employment difference is not statistically significant. The net congestion effect is also an order of magnitude lower than the pair-level treatment effect. At the vacancy-worker pair-level, we find that the matching probability of treated pairs is 5% higher (compared to pairs where neither vacancy, nor workers are treated). This suggests important reallocation effect of recommendations. Averaging workers employment over both types of recommended jobs (control and treated), we find that the net employment effect is smaller and amounts to 2% (p-value=0.10), with again a small negative contribution of congestion effects.

We conduct a thorough heterogeneity analysis of the pair-level employment effects. Those effects are twice as large for less educated and unemployed workers, as well for those who search initially further away from their residence. They are larger when the recommended vacancies have more jobs attached to them, and when the vacancies are recommended longer after their publication. As recommendations are worker-specific, we explore heterogeneous effects along pair dimensions. We compute two distances between recommended jobs and workers' reference job, in terms of geographical distance and occupational distance. The occupational distance is based on actual job-to-job transitions in Swedish administrative data. We find that recommendations that broaden job search in the occupational dimension yield higher treatment effects on matching probability.

This paper contributes to the literature on the effects of Artificial Intelligence technology and machine-learning algorithms on market clearing (Milgrom and Tadelis, 2018). This recent literature mostly considers standard product markets, while we focus on an important matching market, the labor market. We also extend the recent literature documenting the effects of broadband internet on labor markets (Bhuller et al., 2023), or the effects of Craigslist.com (Kroft and Pope, 2014). (Bhuller et al., 2023) and (Kroft and Pope, 2014) find mixed employment effects in the first decade of internet ages, while we study the next generation of online search and matching technology (Kircher, 2022).

Our paper is related to recent experiments recommending occupations to job seekers. In a first lab-in-the-field experiment, Belot et al. (2018) find that recommending occupations that broaden the search of narrow searchers increases their probability to be interviewed. In a larger sample of long-term unemployed workers, Belot et al. (2022) confirm the effectiveness of occupational recommendations, while Altmann et al. (2022) document potential displacement effects of occupational recommendations. We confirm in our setting the effectiveness of occupational broadening down to employment outcomes. Our results tend to downplay the importance of displacement/congestion effects identified thanks to the two-sided randomization (and confirmed using standard clustered/market-level randomization of recommendation treatment).[2] As our recommendations are about specific vacancies (and not occupation) and differ from one worker to another, we are able to investigate new dimensions of heterogeneous effects, wrt geographical search, vacancy popularity, etc. Those dimensions are useful for future design of ML recommender systems on any online job board.

Our paper is also related to the literature analysing the value of information in specific online markets, where not only contacts but all the work relationship remains online. Pallais (2014) uses Upwork to show the importance of feedback information provided by past employers. Horton (2017) shows in the same context that algorithmic recommendations of workers to employers ease the recruitment process. Our results provide evidence on the value of information in at-scale labor markets, namely defined by almost all vacancies posted online in Sweden.[3]

Our analysis relates more broadly to the empirical literature on job search that uses data from online job boards (Marinescu, 2017; Baker and Fradkin, 2017; Marinescu and Rathelot, 2018; Banfi and Villena-Roldan, 2019; Faberman and Kudlyak, 2019; Marinescu and Wolthoff, 2020; Kudlyak et al., 2020; Brown and Matsa, 2020; Hensvik et al., 2021). Our

[2]Accounting explicitly for congestion is possible, as shown by Bied et al. (2023), who propose a recommendation algorithm that maximizes the overall number of matches, at the cost of breaking users' anonymity at the time when the recommendations are generated.

[3]The personnel literature studies how recruiting technologies of individual firms affect hirings (Hoffman et al., 2017). Our paper shows how matching technologies of intermediaries affect the whole labor market.

paper illustrates how matching the online job board data to administrative registers yields important insights on modern labor markets and how they are impacted by technological progress (such as AI).

The paper proceeds as follows. We describe the Swedish institutional background and the data in Section 2, the job recommender system in Section 3, and the RCT design in Section 4. We present the average treatment effects in Section 5. We discuss theoretical considerations in Section 6 to motivate the pair-level analysis of channels from Section 7. We conclude in Section 8.

2 Background and data

The core institution of our analysis is the *Platsbanken.se* platform, the largest online job board in Sweden. Platsbanken is operated by *Arbetförmedlingen*, the Swedish Public Employment Service (PES). On Platsbanken, any private-sector firms or public-sector organizations can post vacancies and screen applicants (free of charge). The coverage of *Platsbanken.se* is very large. According to Eurostat, the average number of vacant jobs in Sweden is 96,569 in 2019Q4. Using the same methodology as the source survey for the Eurostat statistics, we obtain 92,858 job openings in Platsbanken for the same period. The two counts align remarkably well (see also Appendix Figure F1 comparing the industry distribution across sources).

Users can search and view ads and apply to posted vacancies (free of charge). Searching the vacancy listings can be done with free text or by indicating an occupation or a location (see screenshot in Figure 1a). After hitting the search button, users are shown a list of job ads relevant to their criteria (see Figure 1b). The list shows the job ad title, the job location, the employer posting the vacancy and the publication date. To learn more, users can click on a job link and end up on the vacancy webpage. There, users can read the detailed job ad text and other vacancy characteristics (see Figure 1c). To apply for the job, users hit the application button on the top-left of the vacancy webpage.

Our primary data source consists of the records of the online search activity on Platsbanken, combined with the description of all posted job ads. On the vacancy side, the data contain rich information about the posted job, such as the occupation, location, start and end date of publication, working hours, or skill requirements. There is also a firm identifier, which allows us to map each vacancy to firm-level industry codes according to the Swedish SNI classification. On the job seeker side, our data allow us to follow users over time via an anonymized identifier. For each user, we have information about the vacancy id of the viewed ad and a time stamp. An ad view or click is generated every time that a user accesses the vacancy web page via their browser. Users typically end up on the vacancy webpage after clicking on the vacancy list displayed as search results. On top of ad views, we also have information about whether users start the application process for the job.

We also have access to unemployment registers, which provide additional information about unemployed workers that are registered at *Arbetförmedlingen*. The data includes socio-demographics information, such as gender, age, nationality, education level, field of education, and place of residence. We observe the start and end dates of unemployment spells. At the time of registration, unemployed workers report their preferred occupation for which they have required qualification. Occupations are coded into the Standard Swedish Occupation classification (SSYK), similar to the International classification (ISCO) and the US SOC. At the 4-digit level, it has over 400 different occupational categories.

For workers registered as job seekers at *Arbetförmedlingen* (at least once since 2019), we are able to access information from monthly employment registers (from 2019 to 2022). The data include monthly earnings, separately from every employer with their employer id. The employer id allows to match the vacancy data and the employment registers. Therefore, we observe whether a worker applying to a vacancy posted by a specific firm are employed by this firm later on. This allows us to proxy for application success.

Our main sample of analysis consists of workers visiting *Platsbanken.se* between the 1st of April 2021 and the 31st of March 2022, when the job recommender system was live on the platform. We observe their search activity on the website from June 2020 to June 2022.[4] We observe the monthly employment from January 2019 to April 2022 of those workers registered at least once to the Swedish PES over the period January 2019 to June 2022. This sampling scheme implies that we have both employed and unemployed workers over the test period.

3 Job recommender system

In this Section, we describe the job recommender system tested on the Swedish *platsbanken.se* website. The machine-learning algorithm uses ad views (which user views which ad) as input. This kind of data can be considered as *natural-occurring*, in the sense that the data are generated by users on the website during a normal visit. The choice of the input data comes from legal and operational constraints that are common across many online

[4]While online clicks were recorded starting in 2019 (see Hensvik et al. (2020)), the collection of application data started in June 2020 only.

job boards. Namely, using external data from administrative registers (for example to condition recommendations on previous jobs or on demographics) is not feasible. Although we match online search data and employment registers for our analysis, the matching has been approved for research purposes only, and could not be used by the Swedish PES for everyday operational purposes (according to the usual interpretation of the GDPR data protection laws in Europe). While those constraints limit the job recommender system flexibility, they ensure the portability of our recommender system to other websites and ensure the external validity of our evaluation.

3.1 Algorithm

In partnership with the Swedish PES, we build an item-to-item collaborative filtering (CF) recommender system. The objective of the recommender system is to recommend vacancies (items) to job seekers (users).

The recommender system makes use of the implicit feedbacks that job seekers provide when clicking on a vacancy. These feedbacks are stored into a user-item rating matrix R with job seekers i as rows and vacancies j as columns. $R(i,j)$ is the number of times that job seeker i clicked on vacancy j.

These implicit feedbacks are used to estimate the unobserved types of the job seeker and of the vacancies (embeddings). Types are real vectors of length K. The dimensionality of unobserved types is a hyper parameter of the recommender system, set at $K = 128$ in our application. We define the matrix X of job seekers' type where row i contains the types of job seeker i. The dimension of X are (I, K) where I is the total number of job seekers. Similarly, we define the matrix Y of vacancies' types, with dimension (J, K).

The recommender system minimizes the following loss function over unobserved types (X, Y):

$$\mathcal{L} = \sum_{i,j} C(i,j) \left(P(i,j) - X(i,.)Y'(.,j) \right) + \lambda \left(\|X\|^2 + \|Y\|^2 \right) \tag{1}$$

Where $C(i,j)$ and $P(i,j)$ are built from the ratings matrix R: $C = 1 + f(R)$ and $P = sign(R)$. These two matrices help to take into account the implicit nature of the feedbacks. Zeros in the rating matrix are generated by both active ratings - job seekers are aware of the vacancy but did not click it because they find it not suitable -, and by lack of awareness. The matrix C is then a measure of the confidence that the rating is explicit. In the application, $f(.)$ is a cubic function with slope $\alpha = 25$ which is another hyper parameter of the recommender system. The second term of the loss function $\lambda\left(\|X\|^2+\|Y\|^2\right)$ regularizes the optimization with hyper parameter $\lambda = 0.01$.

The above loss function is not a convex problem. We use as an estimation algorithm, a Weighted Alternating Least Square (WALS). The algorithm is described in Hu et al. (2008) and Takács et al. (2011), and we use the Implicit library in Python by Frederickson (2017).

Given estimated unobserved types X and Y, we define the matching score between job seeker i and vacancy j as $\mathcal{M}(i,j) = X(i,.)Y'(.,j)$. For a given job seeker i, we rank vacancies in descending order according to $\mathcal{M}$. Consequently we define the ranking function: $\mathcal{R}(i) = (j_1, j_2, ...)$ where j_1 is the vacancy with the highest matching score for individual i, j_2 the vacancy with the second highest score. The recommender system of rank r returns the r highest ranked vacancy for each individual excluding the history of their clicks:

$$\mathcal{R}^r(i) = (j_1, j_2, ...j_r)/\mathcal{H}(i) \tag{2}$$

where $\mathcal{H}(i)$ is the history of clicks $(j|P(i,j) = 1)$. To increase computation efficiency, we used the NMSLib library to compute recommendations (Naidan et al., 2019).[5]

3.2 Implementation of the recommender system

Given the very nature of the job recommender system, it cannot give recommendations to job seekers without any click history. We search for clicks in the last 30 days before the training day, and we include all users that clicked on at least three different vacancies during this training period.

One risk of collaborative-filtering recommender system is that they recommend to many users the same popular item. In the context of a matching market, this may generate congestion effects. To control that risk, we filter out from the recommendation sets vacancies that received more than 200 clicks over the training period (around 15% of the vacancy population).

To ensure that job recommendations are up to date, the recommender system is trained every day d. The training period is thus a rolling window of 30 days (from $d-30$ to $d-1$). Job recommendations of the d-training vintage are displayed to users during day $d+1$.

In practice, we generate 10 recommendations per user. These recommendations may have become obsolete since they last appeared in the training period, i.e. they are no longer

[5]We use as parameters: $M = 32$; $Post = 1$; $efConstruction = 800$; $ef = 800$

posted on Platsbanken as of day $d + 1$. We exclude these recommendations from the recommended set. Consequently, the number of recommended vacancies may vary across user (and across day).

The *Platsbanken* managing team decided to show job recommendations on the welcome page and on vacancy webpages (see Figure 1c). The first four recommendations are displayed at the bottom of the webpage (see Figure 1c). It is possible to hit a "show more" button to look for subsequent recommendations within the user recommendation set. For each recommended job, the job title, the employer and the job location are listed. The recommended jobs come under the title "Suggestions for you", which emphasizes that the service is personalized. Indeed, let us also emphasize here that each user gets her own list of recommendations.

3.3 Ex-ante precision and coverage

Before presenting the experimental results, we briefly assess the properties of the recommender system. We compute two standard metrics, precision and coverage, used to ex-ante score recommender systems in the machine-learning literature. In our experimental context, we can use the sample of control workers who are not shown the recommendations to estimate those scores. Note that we generate list of recommendations for all workers, whether they are control workers or treated workers who are shown the recommended vacancies.

We first assess the ex-ante relevance of our recommendations by computing the precision of the recommender system. This answers the following question: what is the probability that users click spontaneously on the recommended vacancies? We find that 3% of users click on the highest rank vacancy the day when the recommendations would have been shown. Given the number of available posted ads, the probability to click on a given vacancy at random is about .001%. The recommender system is thus able to pick relevant vacancies to users. In the Appendix, we present a complete analysis of the Mean Average Precision (MAP) score which explores the relevance of recommendations further down in the list.

Even though it is reassuring that our recommendations are ex-ante relevant (and thus will be unlikely to trigger major deception among treated workers), the precision score does not measure the value of recommendations, which will be identified in treatment effects only. Namely, the precision score increases when more popular vacancies are recommended, for which explicit recommendations may not trigger further interest anyway, and application success may be lower because of higher competition.

The second usual metrics of recommender system is coverage. This answers the following question: what is the probability of a given vacancy to be included into at least one recommendation set? We find that 58% of vacancies are recommended to at least one user in a given day.[6] We specifically explore how coverage varies with vacancy interest, proxied here by the average number of clicks per vacancy (over the 30-day training period). The left-hand panel of Figure 2 plots the share of vacancies recommended by click group. The right-hand panel shows the average number of users those jobs are recommended to (for recommended vacancies only). For instance, if during the training month a vacancy is clicked by 60 users, it has 50% chance to be recommended (to at least one user). If this mildly-clicked vacancy is recommended at least once, on average one hundred users receive it as a job recommendation (out of a daily average of 662,392 job seekers). Overall, Figure 2 confirms that vacancies generating more clicks are more likely to be recommended to at least one user and/or to many users.

Overall, these two properties of our job recommender system are well-known in the collaborative-filtering literature. There is a trade-off between precision and coverage.

3.4 Geographical and occupational breadth in recommendations

We complement the usual ex-ante analysis with a comparison between recommended jobs and the jobs that control users spontaneously consider. We answer the following questions: to what extent do recommended jobs differ from jobs usually considered by workers? Is there a scope for the recommender system to broaden workers search or to direct their search to vacancies where they face less competition?

Table 1 reports the characteristics of vacancies that control users click, apply for, and of vacancies posted by firms hiring control users, in Columns (1) to (3) respectively. In Column (4), we report the average characteristics of vacancies in the recommendation set of the same control users, generated for the day when they clicked on the vacancies in Column (1). Column (5) further restricts to top ranked recommendations. In the first row, we consider the geographical distance between the vacancy location and the worker residence.[7] Among clicked jobs, the average distance to the worker residence is 50 km, applied-for jobs are 3km closer, and jobs in which the worker is finally employed are 17 km closer. At every step of the search process, job seekers narrow down their geographical radius. On average,

[6] It is also relevant to consider this statistics over the posting duration. Then XX% of vacancies are recommended at least to one user over their posting duration.

[7] For vacancies, job postings include geographical coordinates of workplace. In the administrative registers, we have the municipality of workers residence, which we locate using the coordinates of its centroid.

recommended jobs are closer to workers than clicked/applied-for jobs, but still further away than where job seekers accept offers.

The next rows in Table 1 report whether the vacancy occupation corresponds to the user reference occupation. For registered user, we consider as reference the occupation of the last job he qualifies for. As a first occupational distance, we rely on the hierarchical structure of the Swedish occupation classification (SSYK) and we consider as more similar occupations that share a greater number of first digits (up to a maximum of four digits). We find that 11% of clicked jobs have exactly the same occupation code as the worker reference occupation (there are over 400 4-digit occupation codes). 15% share the same three first digits, 20% the first two digits and 32% the first digit. We summarize this information into an occupational distance that varies between 0 and 1.[8] We find that average occupational distance decreases within the job search process from clicks to accepted offers as for geographical distance. However, the average occupational distance of recommendations is greater than that of clicked jobs.

We consider another measure of occupational distance based on observed occupational transitions (as in Belot et al. (2018) and follow-up papers). In Swedish administrative data from 2014 to 2018, we follow job-to-job transitions and track occupational changes. Denote τ_{od} the share of transitions to occupation d among all transitions from occupation o. We define as the distance between occupation o and d: $d(o,d) = 1 - \tau_{od}$. We find similar patterns between the various job search steps and the recommendations with that alternative distance. Overall we find that on average recommendations may broaden job search in terms of occupations, but not necessarily in terms of geography. Of course, the broadening effect may depend on whether individual workers are narrow or already broad in their search. We investigate the heterogeneity in ex-ante effects below. Before that, we consider two other characteristics of vacancies: popularity and age (days since publication).

We measure vacancy popularity as the number of daily clicks received by the vacancy during the first 30 days after publication from *control* users. Note that this measure of popularity is not affected by the recommender system as it is computed on control units only and it uses clicks for which indirect spillover effects can be ruled out. It can be considered fully exogenous. On average, workers click on vacancies that 12 control users have clicked per day (in the month following its publication). Applied-for vacancies are slightly more popular, but accepted offers are less so. Workers are probably more likely to receive offers from jobs where there is less competition. Recommendations go in that direction and are

[8]Specifically, we divide by four the number of same first digits between the two occupation code *Sim* and we take as distance $1 - Sim$.

even less popular than accepted offers, opening up opportunities to reduce market congestion. We strengthen that analysis introducing popularity quantiles where we control for difference in tightness across local and occupational markets. We regress the vacancy-level popularity measures on 4-digit occupational dummies and on market fixed effects, where markets are defined as 2-digit occupation X commuting zones X quarter. We sort the regression residuals in quantiles. For that residualized popularity measure, we also find in Table 1 that recommended vacancies are less popular.

Last, we find that workers click on vacancies that are on the website for 12 days on average. Recommended vacancies are significantly older when they are recommended. This is interesting as those older vacancies may be buried at the bottom of the listings shown after a search from the welcome page, and less salient to workers who may have missed them when the vacancies were younger.

We now turn to the heterogeneity of potential ex-ante recommendation effects by workers search types. We consider as search types: (i) whether the user clicks on popular vacancies (i.e that receive above median number of clicks during their first 30 days of publication), (ii) whether her geographical search is narrow (i.e. the average geo distance between clicked jobs and reference job is below across-user median), and (iii) whether her occupational search is narrow (i.e. the average occupational transition-based distance between clicked jobs and reference job is below median).[9] We stack those types in a three-dimensional vector X_i.

First, we consider potential recommendation effects on popularity. We define as a dependent variable the difference between the popularity of recommended jobs and that of clicked jobs, normalized by the standard deviation of the popularity among clicked jobs. We regress the dependent variable on search types X_i, and we report in the upper panel of Figure 3 their coefficients. We find that workers initially clicking on popular vacancies have recommended vacancies one standard deviation less popular.

Second, we consider the difference between the geographical breadth of recommended jobs and of clicked jobs, normalized by the standard deviation of breadth among clicked jobs. In the intermediate panel, we find that narrow searchers in the geographical dimension have recommended jobs broadening the geographical radius of their search.

Similarly, when we consider the normalized difference between the occupational breadth of recommended jobs and of clicked jobs, narrow searchers in terms of occupation have recommended jobs with occupation further away in the occupational space.

Overall, the ex-ante analysis of the recommender system highlights its relevance (limit-

[9]Search types are defined over click activity between April 2021 to March 2022.

ing potential deception among users) and its wide coverage. While coverage is larger for popular vacancies, recommended jobs are still less popular than spontaneous clicks and applications. In addition, recommended jobs tend to broaden workers search in terms of occupation and geography (esp. among narrow searchers for the latter). Those ex-ante properties suggest that the recommender system may spur matching outcomes. We now describe the experimental design to evaluate the ex-post value of the recommender system.

4 Experimental design

After designing the job recommender system, we design a randomized controlled trial in order to evaluate it. The RCT is two-sided, with both users and vacancies being randomized in or out of treatment. From the user side, treated job seekers are shown job recommendations. From the vacancy side, treated vacancies may be shown to users as recommendations, while control vacancies are never shown to any users as recommendations. Over the last five months of the experiment, we add a market-level randomization layer to the user- and vacancy-level designs. After defining local markets by commuting zones and skill level, we randomize half of them into a super control group where vacancies are never shown as recommendations.

4.1 Randomization

The experimental populations are defined according to the job recommender system training. Users for whom recommendations are generated are included in the RCT. Similarly, vacancies appearing in the recommendation list of at least one user (see $\mathcal{R}^r(i)$ definition in previous section) are included in the RCT. The first day a given user is included in the training sample (day d), she is randomized into either the treatment group or the control group with probability 1/2. The treatment status is constant over time. Then treated users will see recommendations from day $d+1$ onwards (until she eventually stops to visit the website, and for at least 30 days).[10]

By comparing treated users and control users, we identify the individual treatment effect under the Stable Unit Treatment Value Assumptions. This is a credible assumption when we consider outcomes without expected spillovers or general equilibrium effects, for example clicks or applications. When we consider job finding rates, SUTVA may be challenged by displacement effects.

[10]The 30-day duration is implied by the rule selecting users in the recommender system training set.

In order to identify the value of the recommender system for individual recruiters, we also randomize the other side of the market. As for users, we randomize vacancies included in the recommendation sets of the experimental population of users either as treated or as control vacancies (with probability 1/2). Treated vacancies will be shown to users if they belong to their recommended sets, whereas control vacancies will never appear on the website as recommendations, even if they belong to some treated users' recommendation sets. The treatment status of vacancies is drawn once and for all, it is constant over time.

Note that recommendations will be effectively shown to treated users only if they visit plastbanken.se after their randomization draw (more precisely after being included in the training/experimental sample). Of course, treatment is ineffective to the treated users until recommendations are shown. They do not receive any specific information about their treatment status and the recommendation services before visiting the welcome or any vacancy webpages. We check that being randomized into treatment has no effect on the probability that users view at least one vacancy over the period when recommendations are generated (see Appendix Table T2). Consequently, we restrict the evaluation sample to those *active* users in the main analysis.

Note also that vacancies in the treatment group will be effectively treated on a given day under two conditions. First, they need to appear in the day-d recommendation set of some treated users. Second, the corresponding treated users need to visit platsbanken.se.

To measure potential displacement effects due to the recommender system, we implemented an extra market-level layer of randomization since November 2021. We aim at isolating some labor markets from the experiment and use them as super controls. We define local labor markets as commuting zones by skill group. We thus split each 69 Swedish commuting zones into two skill groups: High vs. Low. To ensure some balance between treated and supercontrol markets, we perform a paired randomization, from which we excluded Stockholm.[11] In a first step, we cluster local markets into pairs using the number of vacancies and the average number of clicks and of applications per vacancy as matching variables. We report the details of the pairing step in the online Appendix B. Second, within each pair, we assign randomly one market to the super control group and the other remains exposed to the recommender system. From an operational point of view, it is easier to assign vacancies than users to markets, as recruiters declare the workplace location and the job skills as early as when they post the vacancy. To implement the market-level randomization, we exclude from all recommendation sets shown on the website the vacancies

[11] The commuting zone of Stockholm is hardly comparable to any other Swedish commuting zones.

that belong to a super control markets.[12]

4.2 Main evaluation samples and balancing tests

We analyze experimental data from April 1st 2021 to March 31st 2022. We consider two main sample selections. First, we consider any Platsbanken active users and corresponding vacancies over the evaluation period. We denote $\mathcal{S}_a$ and $\mathcal{S}_v$ the respective samples. Second, we consider the subsample of active users who appeared at least once in the unemployment registers from January 2019 to June 2022, denoted $\mathcal{S}_u$. For those users, we have been allowed to match our online search activity dataset to unemployment/employment registers, and thus we have more data. We observe their socio-demographics characteristics and their employment outcomes. For users in the first sample $\mathcal{S}_a$ (but not in the subsample $\mathcal{S}_u$), we only know their history of clicks and applications, together with the characteristics of the vacancy clicked or applied for.

We exclude the supercontrol markets from the main evaluation samples $\mathcal{S}_a$ and $\mathcal{S}_v$ and perform a separate displacement analysis. When we analyze main effects from the recruiters perspective, we exclude since November 2022, all vacancies which belong to super control markets according to their municipality and occupational skill groups. When we analyze main effects from the worker perspective, we exclude since November 2022, users whose reference municipality and reference occupation skill groups belong to a super control market. We define as reference municipality of user i the most frequent municipality of vacancies clicked in the user pre-randomization period. In other words, for every user, we explore her history of clicks in the 30 days before her first randomization day when recommendations are generated for her, and tag the modal municipality and the modal occupation. When users are registered, we prefer to use as references their municipality of residence and the occupation that they state as preferred and which they have qualification for. Those data come from the unemployment register.

The overall sample of active users and corresponding vacancies contain 1.7 million users and 605,000 vacancies. The main evaluation sample of active registered users contains around 245,000 workers. We report in the Appendix the balancing tables for the overall sample of active users and for the sample of vacancies. For active users, we compare across treatment groups the number of clicks on and of applications for vacancies over the 30 days before the randomization. We also compare the characteristics of the vacancy clicked (location, occupation, contract type, hours worked, experience requirement, firm industry).

[12] We show in online Appendix B that this effectively empties the list of recommendations of users whose reference market is in the super control group. Our definition of local markets generates segmentation.

Out of 19 balancing tests in Appendix Table C2, only one is statistically significant at the 5% level. When testing balance in each reference occupation (around 400 categories) and in each reference municipality one by one, we find the expected share of 95% non-rejected tests at the 5% level (see Appendix Table C1). Similarly, characteristics of experimental vacancies are balanced across experimental arms (see Appendix Tables C5 and C6). Interestingly, vacancies are on *Platbanken.se* for on average 4 days before being included in the recommender system. Treatment occurs relatively early in the vacancy lifecycle as the median application deadline is 29 days after publication (see Appendix Figure F6 for the distribution of time to deadline).Vacancies receive around 70 clicks and 5.5 applications before randomization. We report in the Appendix the evolution of application received per week since publication for the inflow of control vacancies (see Appendix Figures F7 and F8). On average, control vacancies receive 5.4 applications per week which compares well with other estimates in the literature (see Appendix Table T1). For example, Marinescu (2017) finds that vacancies posted on CareerBuilder.com receive 7.5 applications per week, and Banfi and Villena-Roldan (2019) finds that 4.06 applications are recieved by vacancies posted on trabajando.com.

We also check in the Appendix the balance among control vacancies between those in super control markets and those in markets with treated vacancies.

We now focus on the sample of registered users where socio-demographics and employment history are available. Table 2 checks the balance of pre-randomization covariates across treatment-control groups. None of the difference between treated and control means is statistically significant at the 5% level. In the evaluation sample $\mathcal{S}_u$, one out of two unemployed is a woman, around 45% are not Swedish, 10% live in Stockholm (see the Appendix Figure F5 for the other most frequent municipalities). One quarter are high school dropouts, and around one third of unemployed have a post-secondary diploma. In Appendix Figure F4, we list the most frequent occupations at the 2-digit level: 13% look for personal care jobs, 9% for sales jobs, other occupations make up less than 5% each. We verify balance across experimental arms for categorical variables (occupation and municipality) in the online Appendix Table C4. Before the randomization month, and since January 2019, their average monthly earnings are around 1,007 euros (gross). As they are employed 40% of the months over that period, this yields average monthly wages at 2,067 euros. During the month before randomization, workers visit *Platsbanken.se* on average 3 days, cumulating 14 clicks on vacancies and making 4 applications. These job search statistics compare well with other estimates from the literature. Faberman and Kudlyak (2016) document that job seekers on the platform SnagAJob.com apply to around 8 jobs per month through the website. Marinescu and Rathelot (2018) provide a comparable estimate of 4.3 applications per month through CareerBuilder.com. Faberman et al. (2017) find in the Job Search Supplement of the Survey on Consumer Expectations that US job seekers apply to 4 to 8 jobs per month depending on their employment status, whatever the application media (through web platforms, physical contacts, etc.).[13]

5 Impact of recommendations from workers' perspective and from recruiters' perspective

We first estimate the treatment effects on job search and matching outcomes from the worker perspective and from the vacancy perspective independently.

5.1 Workers' perspective

We estimate the treatment effects from the worker perspective. We first conduct a daily worker-level analysis on the largest sample of active users ($\mathcal{S}_a$). We collapse the click and application data at the worker level for any given visit day when recommendations are generated. We run the following regression:

$$Y_{id} = \alpha + \delta T_i^u + \varepsilon_{id} \tag{3}$$

where Y_{id} is the outcome for user i during day d, and T_i^u is the treatment status of user i. We cluster the standard errors at the user level.

Table 3 reports the treatment effect on daily clicks on different vacancies in Panel A and on daily applications in Panel B. In Column (1), we count clicks and applications on any jobs whether recommended or not. In the upper panel, we find no statistically significant treatment effect, and we can rule out effect larger than 0.25% wrt the average number of daily clicks for control users. This despite an increase in daily clicks on jobs in the personalized recommendation set of user i that are shown on the website (randomized as treated vacancies). Treated users click more on recommended jobs than control users (Column 2). This represents a statistically significant increase of 44%. For the control group, the recommended vacancies are not displayed as personalized suggestions. Thus the control average is also a measure of the recommender precision. Users make 0.1 daily click on

[13] It is not relevant to compare our estimates with those of Marinescu and Skandalis (2020) (0.3 application per month) who focus on applications *registered* by the French PES. Similarly, Altmann et al. (2022) do not analyze the number of applications per user in their data that are restricted to registered applications sent by job seekers to their caseworkers from the Danish PES. The information in registered applications is shaped by Danish UI rules that require at least two applications per week.

recommended vacancies without any intervention. This is to be compared to the control mean in Column (1). Recommended vacancies generate up to 3% of clicks without any intervention (=0.105/3.347). In Column (3), we consider the subset of recommended vacancies to user i that are randomized into the control group and are not shown in the website recommendation box. Treatment effects are negative, highlighting a substitution effect (of about 1% of the control mean). In Column (4), we count daily clicks on vacancies that do not belong to the recommendation set of user i. Again, treatment effects are negative, highlighting a substitution effect of similar magnitude in percentage (1%). Users substitute non-recommended jobs for recommended jobs.

Clicks are a first measure of job search intensity, but do not necessarily capture the quality of the recommender system. Users may click on recommended jobs out of curiosity, but may consider them irrelevant after reading the job ads. This would generate positive treatment effects on clicks, that miss some irrelevance issue. To capture the quality dimension, we consider as outcomes job applications in Panel B. Job applications measure users' genuine interest in the job ads (compared to simple clicks). In Column (2), we find a positive treatment effect on daily applications for recommended jobs (by 30%), while the substitution effect observed in clicks persists to the application stage and we obtain a negative treatment effect on applications for non-recommended jobs. Overall the positive effect is overturned by the substitution effects, so that we find a statistically significant decrease in total applications of 0.9%.

The control means in both panels of Table 3 allow to compute conversion rates of clicks into application. For all jobs, the conversion rate of control users amounts to 11.9% (=0.40/3.35). Still for control users, the conversion rate is higher for recommended jobs (13.6%) than for non recommended jobs (11.9% in column 4). This confirms that recommended jobs are positively selected in terms of applications. This is an interesting result as the recommender system is trained with click data only and does not take application as input. Another interesting pattern emerges when we compare the conversion rate on recommended jobs for treated and control users. Indeed it is lower for treated users. Let us assume some monotonicity in search behavior, where individual recommended vacancies that control users click on would be clicked had users being treated and shown the list of recommended jobs. Under that assumption, the lower conversion rate in recommended jobs for treated users suggests that the marginal recommended vacancy clicked by treated users are slightly less attractive. This may be explained by the vacancy itself or by some timing issues. In Table 3, we implicitly compute within-day conversion rates, which assumes only short delay between viewing and applying for the vacancy. For marginal clicked vacancies that come as a surprise to the job seeker, it seems reasonable that it takes some time to prepare the application package and to apply later than the day when the vacancy is recommended. In the pair-level analysis below, we relax this time constraint when analyzing application.

In Table 4, we estimate treatment effects on reemployment outcomes. Those outcomes are observed in the sample of registered users only.[14] For every registered user i, we tag the first day when recommendations are generated within the experimental period and the corresponding calendar month. We then consider three different outcomes observed in the monthly employment register after the randomization month until June 2022 (the last observed month in our dataset). We consider whether the user received any earnings over that period (Column 1), the average monthly earnings (Column 2, including zeros in months when users are not employed) and the fraction of months with positive earnings (Column 3). We run the worker-level regression of the reemployment outcomes on a treatment dummy and report the treatment coefficient in Table 4. In Panel A, we do not control for any worker covariates. In Panel B, we include all the worker covariates of the balancing analysis. In Panel C, we let the double-debiased machine learning estimator select the relevant set of covariates and interactions. Overall, we find a positive impact on employment, statistically significant at the 5% level in Panels A and B and at the 10% level in Panel C. From a baseline reemployment of 60.5%, employment increases by 0.3 to 0.4 percentage point, which represents a 0.5% to 0.7% increase. .

In Columns (2) and (3), we find that the treatment effect is positive, although statistically significant in Column (3) of Panel A only. The percentage impact is of the same order of magnitude across columns (around 0.5%).

Overall, we find that treated users shift their search intensity towards recommended vacancies, leading to a slight increase in employment. We do not find any significant increase in match quality (proxied by employment duration). On average, we observe employment over the 6 months after randomization, which may be a short horizon to capture match quality effects.

5.2 Recruiters' perspective

We estimate the treatment effects from the recruiter perspective. We first conduct a daily vacancy-level analysis. For every experimental vacancy, we include in the regression all days when it is included in the recommendation sets. There may be endogenous selection in the sample, as vacancies with positive treatment effects become more popular and

[14]We check in the Appendix that the treatment effects on search activity (clicks and applications) are of similar magnitudes on the subsample of registered users (see Appendix Table T4).

may appear more frequently in recommendation sets. Appendix Table T3 shows that treatment does not correlate with the number of days vacancies appear in at least one recommendation set, which supports our daily sample construction. We collapse the click and application data at the vacancy level, and we run the following regression:

$$Y_{jd} = \alpha + \delta T_j^v + \varepsilon_{jd} \tag{4}$$

where Y_{jd} is the outcome for vacancy j during day d, and T_j^v is the treatment status of vacancy j. We cluster the standard errors at the vacancy level. We report the δs coefficients in Table 5, where the upper panel corresponds to clicks received while the lower panel corresponds to applications. In Column (1), we count clicks received from any users (within or out of the experimental population). Treated vacancies receive significantly more clicks. On average, control vacancies receive 8.5 clicks per day, this increases by 1.1% when vacancies are recommended to users. This overall effect is driven by marginal clicks by users who have that specific vacancy in their personalized recommendation set (see Column 2), and more specifically treated users who will be shown the recommendation box during their visit (see Column 3). Treated vacancies receive 48.8% more clicks from treated users. There is no reason why treated vacancies should receive more clicks from control users and indeed we find a small coefficient in Column (4).

In the lower panel of Table 5, we find similar patterns for applications. Treated vacancies receive more applications than control vacancies. Specifically they receive 30% more applications from treated users for whom that vacancy is recommended. As for the worker-level analysis, the split across columns requires that vacancies appear in the recommendation set within the same day, which is a strong condition when analysing applications. Search outcomes may require some delay after the vacancies are viewed in the recommendation box. In the pair-analysis below, we relax that tight timing condition.

Do the marginal applications on treated vacancies lead to more hires and higher firm growth? We investigate those effects in the sample of firms posting at least one experimental vacancy. For every firm, we compute the share of vacancies in the treatment group over the whole experimental period ($ShareT_f^v$). We leverage the panel structure of the matched employer employee registers and we compute monthly hiring rates and employment growth rates for all firms from January 2019 to April 2022. The monthly treatment effects are obtained from the following regression:

$$Y_{f\tau} = \sum_{\tau=\text{Jan 19}}^{\text{Apr 22}} \alpha_\tau + \delta_\tau ShareT_f^v + \varepsilon_{f\tau} \tag{5}$$

where $Y_{f\tau}$ is the outcome of firm f in month τ. The left-hand side panel in Figure 4 reports the estimated coefficients δ_τ together with their 95% confidence interval for monthly hiring rates, while the right-hand side panel considers monthly employment growth rates. The vertical line indicates the first month of the experimental period April 2021. Before that date, the δ_τ coefficients are placebos/balancing tests. After that, they identify treatment effects. Overall, the δ_τ coefficients are small and not significant. This suggests that marginal applications due to the recommender system do not trigger significant effects on firm employment.

Alternatively, our setting may be underpowered to detect firm-level effects. First, while significant, the treatment effect on vacancy-level application is 2% which requires a strong elasticity of hirings to applications to trigger significant effects on monthly rates. Moreover, due to operational constraints, the demand side of the market is randomized at the vacancy-level, which dilutes treatment differences across firms as they post several vacancies. The Appendix Figure F9 shows the distribution of the treatment share across firms and how almost 15% of firms has a treatment share of 50%.

5.3 Displacement effects

In the two previous analysis, we document relative effects for treated individual units compared to control individual units. Those individual treatment effects identify policy-relevant effects under the SUTVA assumption and in the absence of spillovers, or general equilibrium effects. As already stated, this is a credible assumption when analyzing clicks and application behavior from workers' perspective. There are no clear mechanisms that would make clicks or applications of treated workers crowd out those of control workers.[15] However, from that same worker-level analysis, we have learnt that treated workers substituted their applications from control job ads to treated ones. From the vacancy-level analysis, we have learnt that control job ads received less applications: treated vacancies may displace those in the control group. We investigate the extent of those displacement effects leveraging our market-level randomization. We compare daily clicks received by control vacancies posted between December 2021 and March 2022 in markets where 50% of vacancies are treated vs in super-control markets where no vacancies are treated. We run

[15]One potential mechanism could go through vacancy posting, if application of treated users make some specific vacancies disappear from the website at a faster rate preventing control users to click/apply for them later in the spell. We do not observe any treatment effects on the duration vacancies remain available on the website. This is not surprising as the application deadline is set ex-ante by recruiters before their job ad comes live on the website, and there is a strong norm towards a default duration of 1 month as can be seen in Appendix Figure F6.

the following regression:

$$Y_{jd} = \alpha + \delta SuperT^{v}_{m(j)} + RandPairFE + \varepsilon_{jd} \tag{6}$$

where Y_{jd} is outcome of vacancy j for day d. Vacancy j belongs to local labor market $m(j)$ which is randomized in either super treated status ($SuperT^{v}_{m(j)} = 1$) or super control ($SuperT^{v}_{m(j)} = 0$). As randomization is blocked into pairs, we also include fixed effects for market pairs of randomization. The estimation sample does not include Stockholm which is left out from the market-level randomization. We cluster standard errors at the market level. We report the estimates for δ in Columns (1) to (3) of Table 6. We find negative point estimates on daily clicks and applications received, consistent with negative spillover effects. However they are not statistically significant. This leaves little support for important spillover effects, which is further confirmed by estimates on firm size. In Column (3), we select firms posting control vacancies and estimate the super-treatment effect on their log number of employees in June 2022. The coefficient estimate is not statistically different from zero, and if anything it is positive.

6 Theoretical considerations

In the following section, we leverage the personalization of the recommendation list to study worker-vacancy pair-level outcomes. Before presenting the results, it is useful to describe expected effects from a theoretical perspective and what the two-sided randomization allows us to identify.

In the pair analysis, we append to all experimental users the vacancies from their recommendation sets that are both treated and control vacancies. We observe workers' application behaviors and employment outcomes in four different cells defined by the combination of worker and the vacancy treatment statuses (T^{u}_{i}, T^{v}_{j}). We consider expected effects on applications first, and turn to employment outcomes next.

Job search diplacement. First, we compare how control workers apply to control vs. treated job ads. For control users, nothing distinguishes the control and treated ads that belong to their recommendation sets, as they do not see any recommendation. Building up on Rubin's potential outcome framework, we denote $A_{ij}(T^{u}_{i}, T^{v}_{j})$ the potential application outcomes when the joint treatment status of user i and vacancy j is (T^{u}_{i}, T^{v}_{j}). We assume that a job ad's treatment status is irrelevant to control workers' application behaviour:

$$\forall (i, j) \text{ such that } j \in \mathcal{R}(i), A_{ij}(T^{u}_{i} = 0, T^{v}_{j} = 1) = A_{ij}(T^{u}_{i} = 0, T^{v}_{j} = 0),$$

where $\mathcal{R}(i)$ is the recommendation list of worker i. To simplify notation, we define $A_{CC} \doteq E[A_{ij}|T_i^u = 0, T_j^v = 0]$ and $A_{CT} \doteq E[A_{ij}|T_i^u = 0, T_j^v = 1]$, where expectations are implicitly taken over all worker-job (i, j) pairs such that $j \in \mathcal{R}(i)$. Note that, in notation A_{CC}, the first subscript relates to workers' treatment status and the second one to job ads' treatment status. An implication of the previous assumption is that the application rate in both cells with control users is the same: $A_{CC} = A_{CT} \doteq A_0$.

We now consider treated users. When users have a limited attention span, a positive marginal application cost, or a decreasing marginal application return, they may substitute away from control vacancies, as the marginal cost of applications for treated vacancies decreases (or the perceived value attached to recommended jobs increases). In Table 4, we find that treated users are less likely to apply to control vacancies than control users. We formulate the following monotonicity assumption:

$$\forall (i, j) \text{ such that } j \in \mathcal{R}(i), A_{ij}(T_i^u = 1, T_j^v = 0) \leq A_{ij}(T_i^u = 0, T_j^v = 0).$$

This assumption would be satisfied in an application model where users apply to vacancies whenever expected payoff net of marginal cost are above a threshold value, the job recommender system affects user application behavior by increasing the threshold value, even without affecting expected payoffs or costs. The monotonicity assumption implies that $E[A_{ij}|T_i^u = 1, T_j^v = 0] \leq E[A_{ij}|T_i^u = 0, T_j^v = 0]$. We denote the difference between the two terms: $A_1 = A_{CC} - A_{TC} = A_0 - A_{TC}$.

The negative indirect substitution effect on control job ads is the counterpart of the positive direct effect on treated ads. Assuming pair-level monotonicity, we expect:

$$\forall (i, j) \text{ such that } j \in \mathcal{R}(i), A_{ij}(T_i^u = 1, T_j^v = 1) \geq A_{ij}(T_i^u = 0, T_j^v = 1).$$

This assumption implies that $E[A_{ij}|T_i^u = 1, T_j^v = 1] \geq E[A_{ij}|T_i^u = 0, T_j^v = 1]$. We denote the difference $A_2 = A_{TT} - A_{CT} = A_{TT} - A_0$. We summarise these theoretical predictions in Figure 5.

Thanks to our two-sided randomization plan, we have access to the empirical counterparts of A_{CC}, A_{CT}, A_{TC} and A_{TT}. We can identify each component A_1 and A_2, which contribute to total effect on application $A_2 - A_1$ and characterize substitution effects.

Congestion. Let us go back to the subsample of control users. Given that their application behavior is unaffected, any change in employment outcomes is related to competition/congestion effects due to the presence of treated users, as they change their application behavior. As treated users substitute away from control vacancies, control users will face lower competition when applying to control vacancies and greater competition when applying to treated vacancies. We expect

$$\forall(i,j) \text{ such that } j \in \mathcal{R}(i), E_{ij}(T_i^u = 0, T_j^v = 0) \geq E_{ij}(T_i^u = 0, T_j^v = 1).$$

Let us denote $E_0 = E_{CC} = E[E_{ij}|T_i^u = 0, T_j^v = 0]$ and $E_{CT} = E[E_{ij}|T_i^u = 0, T_j^v = 1]$. A net measure of congestion effects is then $E_{CC} - E_{CT} = E_0 - (1-g)E_0 = gE_0$ where we assume $g \in (0,1)$. We note that the presence of treated users may lead to an increase in E_0 (compared to a counterfactual of no recommendations to any user), as there is lower competition on control jobs.

Excluding some vacancies from the recommendation lists allows to identify congestion effects in a conservative way. If we had not done so, the personalization of recommendations implies that any vacancy would eventually be recommended to some users. Even if there had been a subsample of vacancies not recommended to any user, application rates of control users would differ across recommended and not recommended vacancies and congestion effects would be confounded by heterogeneity in conversion rates across recommended and not-recommended vacancies.

For treated users, employment in control jobs decreases, as they apply less, and give up A_1 applications compared to control users. We denote the corresponding employment gap: $E_1 = E_0 - E_{TC}$ where $E_{TC} = E[E_{ij}|T_i^u = 1, T_j^v = 0]$. At this stage, it is useful to distinguish always-applied-for vacancies from substituted vacancies. The employment (conversion) rate on always-applied-for vacancies, i.e. vacancies such that $A_{ij}(0,0) = 1 \& A_{ij}(1,0) = 1$, is E_{TC}/A_{TC}, which can differ from the employment rate on substituted vacancies E_1/A_1 (substituted vacancies are such that $A_{ij}(0,0) = 1 \& A_{ij}(1,0) = 0$). Economic theory would predict that workers give away applications where they have lower chances: $E_1/A_1 \leq E_{TC}/A_{TC}$.

On treated vacancies, treated users increase their applications by A_2, pushing up their average employment. On the other hand, their average employment is pushed downwards as they face greater competition on each treated vacancies. Again it is useful to distinguish treated vacancies, for which users would have applied, had they been control. Formally, those treated vacancy are defined as $A_{ij}(0,1) = 1 \& A_{ij}(1,1) = 1$ and we have the complement treated vacancy type: $A_{ij}(0,1) = 0 \& A_{ij}(1,1) = 1$. In words, these are the always-applied-for vacancies and the marginal vacancies. Then we can decompose employment for treated users in treated vacancies into two terms. The first term corresponds to employment from always-applied-for control vacancies $E_{CT} = (1-g)E_0$, the second term relates to employment from marginal vacancies that we denote $(1-g)E_2$ where the factor

$(1-g)$ clarifies that congestion effects also hit those treated vacancies (with same intensity by assumption). To sum up, we have: $E_{TT} = E[E_{ij}|T_i^u = 1, T_j^v = 1] = (1-g)(E_0 + E_2)$.

7 Worker-job pair design and heterogeneous effects

We present the results from the pair-level analysis. We first compute the net employment effect within recommended jobs and identify net congestion effects thanks to the two-sided randomization. Together with our large sample size, recommendation personalization allows us to conduct a detail and thorough study of the heterogeneity of the recommendation effects, that we present last.

7.1 The worker-job pair sample

We start from the subsample of registered users S_u. We append all recommended vacancies j, such that job ad j belongs to worker i's recommendation set on a day when worker i is active on *Platsbanken.se* website. We obtain a sample of 59 millions of worker-ad pairs.

For each worker-vacancy pair, we tag the first date when vacancy j appears in the recommendation set of worker i. From this date onward (and until the end of our search activity dataset in June 2022), we sum all clicks and applications of worker i on vacancy j. This relaxes the within-day timing assumptions of the previous sections. We define pair-level employment if worker i has some positive earnings in the firm f that posted job ad j after the month when worker i applied for vacancy j. As there are no vacancy identifier in the employment registers, we rely on the application information and employment dates to link worker-firm-level employment spells to job ads. For each employment spell, we record their duration, and total earnings.

7.2 Pair-level effects

We estimate at the pair-level the following regression:

$$Y_{ij} = \alpha_{00}(1 - T_i^u)(1 - T_i^v) + \alpha_{10}T_i^u(1 - T_i^v) + \alpha_{01}(1 - T_i^u)T_i^v + \alpha_{11}T_i^u T_i^v + \nu_{ij} \qquad (7)$$

where Y_{ij} is a pair-level outcome for user i and vacancy j. The regression coefficients yield the expectations of Y on each of the four cells defined by both treatment statuses.

In Panel A of Table 7, we report the αs coefficients for various outcomes across columns. In Panel B, we report the pair-level effects, which we define as the difference between the mean outcomes of treated pairs (with treated jobs and workers) and that of control pairs (with both control jobs and workers): $\alpha_{11} - \alpha_{00}$.

In Columns (1) and (2), we find a decrease in the clicks and applications of treated users to control vacancy and an increase towards treated job ads. The treatment effect estimates are consistent with results from Section 3: 39% increase in clicks towards treated recommended jobs and 21% increase in applications.

Employment effects. Column (3) of Table 7 reports the treatment effects on pair-level employment. First, we find that employment of treated users in treated vacancies is higher than employment of control users in control jobs by 5.5%. This points to an sizable reallocation effect of the recommender system. The control employment mean is low (0.046%) because the application probability is around 1%. However, the probability of hiring conditional on applying, computed as the ratio between the two previous statistics is 4.6% in line with the order of magnitude found in other studies. In Column (4), we find similar pair-level effect of the recommended system on employment duration (3%), although not statistically significant. As in Table 4, the pair-level effect on earnings is not statistically significant.

Combining the coefficients from Column (3) of Table 7, we can reconstruct the treatment effect on employment of workers in treated jobs, in control jobs and in all jobs. We report the estimates of these quantities in the first three rows of Table 8. Recommendations increase the probability of employment on treated jobs by a statistically significant 3.5%. Consistently with the existence of a reallocation of applications from control to treated ones, recommendations tend to reduce the probability to be employed in control jobs by -1.5% (p-value=.075). The sum of the previous two effects is a positive 2-percent effect (p-value=.10).

Congestion. Our design allows us to compare the employment probability of controls workers in treated vs. control jobs. We first check in Table 7 that the probability to apply to a treated or a control job is exactly the same for control workers (Column (2), third and fourth rows). However, we see that the employment probability is lower in treated than in control jobs. In Table 8, we report in the fourth row the net congestion effect g, which corresponds to the share of employment lost because of the differential congestion between control and treated jobs. The point estimate is 1.4%, with a standard error of 1.7%. Its order of magnitude is also smaller than the pair-level effect on employment of 5%.

While the contribution of congestion to net employment effect is small, it is worth discussing whether the net congestion effect (of 1.4%) itself is consistent with our other estimates. As a benchmark, we predict the congestion effects on employment from estimated effects on applications and a simple recruitment model. The employment probability in a given job (match probability) for control workers is the product of application probability (denoted A_{CC} and A_{CT} on control and treated job resp.) and the probability of accepting an offer conditional on applying. We assume that any offer is accepted, recognizing that the probability of multiple offers from the worker perspective is negligible. We thus denote O_{CC} and O_{CT} the probability of accepting/receiving an offer from a given job conditional on applying to that job. We assume that workers are homogeneous, so that every applicant worker has the same chance to receive an offer, and the firm needs to send only one offer to recruit. Then the offer probability is the inverse of the number of applicants to the job: $O_{CC} = 1/E[\sum_i A_{ij}|T_j = 0]$ and $O_{CT} = 1/E[\sum_i A_{ij}|T_j = 1]$. The total number of applicants to a given job can be related to the number of daily applications $E[\sum_i A_{ij}|T_j = t, day = d]$, estimated in Table 5, using the average number of days that vacancies can be applied for from the same table, denoted $N_{days,C}$ and $N_{days,T}$. Moreover, we have that $N_{days,C} = N_{days,T}$ from the Appendix Table T3. We thus write the predicted net congestion effect as:

$$\frac{E_{CC} - E_{CT}}{E_{CC}} = \frac{E[\sum_i A_{ij}|T_j = 1, day = d] - E[\sum_i A_{ij}|T_j = 0, day = d]}{E[\sum_i A_{ij}|T_j = 1, day = d]},$$

This shows that the net congestion effect should be of same order of magnitude than the relative treatment effect on application received by vacancies, estimated at 1.12% in Table 5.

7.3 Heterogeneous effects

The pair-level design and the personalization of recommendations allow to study the heterogeneity of treatment effects by job characteristics, by worker characteristics, and more importantly by worker-job pair characteristics. It allows to finely identify the personalized recommendations that yield the largest pair-level treatment effects.

In this section, we focus on the heterogeneity of the pair-level effects. We work on the subsample of the pair-level data with either pairs of treated workers and treated jobs, or pairs of control workers and control jobs, a subsample of 29.6 million pairs. We estimate the following regression:

$$Y_{ij} = \sum_{l} \sum_{k \in \mathcal{K}_l} \delta^{l,k} \mathbb{1}[X^l_{ij} = k] T^p_{ij} + \sum_{l} \sum_{k \in \mathcal{K}_l} \beta^{l,k} \mathbb{1}[X^l_{ij} = k] + \varepsilon_{ij} \tag{8}$$

where T^p_{ij} indicates whether pair (i, j) is treated, and $\mathbb{1}[X^l_{ij} = k]$ is a dummy indicating whether covariate X^l measured at the level of pair (i, j) is equal to k. In terms of outcomes, we will mainly consider applications and employment (as defined above in this section). In what follows, we report a rescaled version of δ^k, as we are interested in whether recommendations increase (or decrease) each category of pairs' employment probability by different factors, rather than in raw differentials in employment. For instance, if women have higher baseline employment probabilities than men, δ_{women} may be higher than δ_{men} while the percent impact of the treatment is the same. For each variable X^l, there is a category of reference k_0. For all categories, we report the average employment effect for this category. For categories $k \neq k_0$, we also report a 95% confidence interval that tests the difference (in percent impact) between k and k_0. Throughout the analysis, standard errors are clustered by job and worker. It is also worth stressing that all coefficients $\delta^{l,k}$ are obtained from the same regression, meaning that we account for the correlation between the covariates X^l.

Worker-level heterogeneity In Figure 6, we report the treatment effects (δ^k) by worker-level covariates. We plot in each panel the effects on applications (in red) and on employment (in dark blue). In the upper-left panel, we find no significant heterogeneous effects between male workers (on the left-hand side) and female workers (on the right-hand side). This contrast with Behaghel et al. (2022) who find strong heterogeneity by gender. We find that older workers tend to have lower treatment effects on applications but the point estimates for employment are very similar (and clearly not significantly different). In the next two panels, differences in treatment effects are starker. The recommender system produces large effects for workers who dropped out from high school and for unemployed workers at the beginning of treatment. Recommendations increase employment of high-school dropouts by 20% and employment of unemployed workers by about 12%. This is in line with recent results of the Belot et al. (2022) experiment recommending occupations to long-term unemployed.

In the bottom two panels, we leverage our unique data on search activity to test whether recommender systems have differential effects depending on how workers searched before being showed recommendations. We characterize pre-experimental search based on the clicks we observe before workers are randomized in the experiment. For all pre-experimental clicks, we compute the distance between the worker residence and the municipality of the workplace and take the average. We then split the workers' sample based on quartiles in average residence-to-workplace distance. We find that the effects of recommendations tend to increase with initial geographical search breadth. Workers who are initially broader in their search before seeing recommendations benefit the most from the recommender system. In the last panel, we investigate the heterogeneity with respect to the occupational breadth of job search before randomization. Using the clicks prior to randomization, we characterize occupational search breadth by computing the share of clicked jobs with the same 4-digit occupation code as the reference occupation stated by workers to the PES. We split the sample between below and above the population median, we find little heterogeneity in this dimension. This result is in contrast with Belot et al. (2018) who find stronger effects of occupational advice to narrow searchers. One explanation lies in the type of advice generated by our recommender system. We explore in the last section below whether the recommender system yields larger treatment effects with broader recommendations.

Heterogeneity in the kind of jobs recommended. In Figure 7, we report the treatment effects (δ^k) by job-level covariates. Recommendations do not have an heterogeneous effect on the probability to apply depending on whether several jobs (or just one) are attached to a job ad. However, recommending job ads with several jobs has a significantly higher effect on employment than recommending those with just one job. This could be explained by the fact that the number of jobs attached to the job ad is not very salient on Platsbanken, and is not fully taken into account by workers when deciding whether to apply.

We find that job ads recommended to more users trigger more interest and generate larger effects on application. However this does not translate into larger employment effects. In the lower panel of Figure 7, we inspect heterogeneity by vacancy popularity. To compute popularity, we count the number of applications that the job ad received from users in the control group. We then regress this measure on occupation, municipality and year-quarter fixed effects, and compute quintiles of the regression residuals. We find that treatment effects on applications decrease with popularity (from 40% for least popular vacancies to 20% for the most popular). The pattern is less clear for employment. If anything, medium-popularity jobs are those with lower employment, but differences are neither strong nor very significant.

Pair-level heterogeneity In Figure 8, we study the heterogeneity of the treatment effects across dimensions that vary at the ad-worker pair level. We first consider whether the duration since a vacancy is out at the time of recommendation matters. In the upper left-hand panel, we find that time since posted leads to larger effects on applications. More recent vacancies usually appear high in the list after users hit the search button on the welcome page, and we find a negative duration dependence in applications received as a function of vacancy age (see appendix Figure F7). Consequently, larger effects in applications are due to lower baseline applications rates on older vacancies. The heterogeneity is also strong for employment effects: recommending jobs that have been posted more than a month increases employment by 20%, statistically higher than recommending younger jobs. Second, we find that job ads that are ranked higher in the recommendation list have larger effects on applications. This is likely due to a difference in salience, as only the five top recommended vacancies are displayed in the recommendation box by default and users need to hit the "display more" button to inspect the next recommended ones. However, we do no find significant heterogeneity in effects on employment.

Last, we consider heterogeneity in the matching distance between the supply and demand side of the market. First, we compute the geographical distance between the workers municipality of residence and the vacancy municipality, and split the variable in quartiles. We find that the effect on applications does not depend much on potential commuting distance. However, recommending jobs that are further tend to produce lower effects on employment, although these differences are neither significant not systematic. Second, we compute the proximity between the occupation of the recommended job and the users' reference occupation. We distinguish three categories: (i) both occupations have the exact same 4-digit code, (ii) they are related according to the transition-based approach adopted by Belot et al. (2018), and (iii) all other (i.e., further away) occupation pairs. We do find larger treatment effects on applications when recommended occupations are further away from workers' reference occupation, in line with Belot et al. (2018). Results also suggest that employment effects are higher when we recommend jobs that are further away in the occupation space, with significance levels that vary between 5% and 10%.

8 Conclusion

Until now, research on job search assistance has mainly focused on labor intensive forms of assistance (like counselling) or on algorithmic – but non-personalized – occupational advice. In contrast, this paper studies if and how individualised recommendations generated by AI technology can enhance the job matching process on online job boards.

More specifically, we design a machine-learning job recommender system and evaluate it using a large scale clustered two-sided randomized controlled trial on Sweden's largest online job board. Our recommender system uses naturally occurring data on user-level vacancy clicks as input, which makes it transferable to most other online job boards.

We show that the recommender system has several properties that may enhance matching outcomes: it proposes relevant jobs and the recommendations broaden workers search in terms of occupation and geography, especially among job seekers with a historically narrow search radius. Additionally, recommended job ads are less popular than those workers spontaneously apply to and the recommender system tends to increase the salience of older vacancies.

Our evaluation of the ex-post value of the recommender system shows that it had a clear effect on search behaviour: treated job seekers increased their daily clicks on recommended jobs by 44%, while they decrease their clicks for non-recommended vacancies by 1% resulting in a zero overall impact on the number of clicks. Treated job seekers also reallocated their applications from non-recommended to recommended ones, but reduced their total number of applications by around 1%. Treated workers tend to be more likely to be employed after being exposed to the recommender system: the effect is small, around .6%, but the very low marginal cost of the intervention makes any gain worth cost-effective.

When analysing the recommendation effects at the pair level, we find that the matching probability of treated pairs is 5% higher. This highlights the importance of reallocation effects of treated workers towards recommended vacancies. These effects are substantially larger for unemployed and less-educated job seekers.

Importantly, the potential congestion effects of the recommendations appear to be small. This result differs from the findings by Altmann et al. (2022) who document significant displacement effects of occupational recommendations in the Danish context. An important difference to their setting is the personalised nature of our recommendations, which should reduce the negative spill-overs that may arise from coarser occupational advice. We do however conclude that the employment effects are larger when job seekers receive recommendations for less popular vacancies and hence when there is more scope for marginal applicants to get hired on the recommended job.

Together our findings provide strong support for artificial intelligence as a tool to be leveraged on online job boards. As such personalized advice can be scaled up easily and at low cost, future research should continue to explore the properties and features of an efficient job recommender system and if the insights from our study can be extended to other types of matching markets.

References

Acemoglu, D., D. Autor, J. Hazell, and P. Restrepo (2022): "Artificial Intelligence and Jobs: Evidence from Online Vacancies," *Journal of Labor Economics*, 40, S293–S340.

Agrawal, A., J. Gans, and A. Goldfarb (2019): *The Economics of Artificial Intelligence: An Agenda*, University of Chicago Press.

Altmann, S., A. Glenny, R. Mahlstedt, and A. Sebald (2022): "The Direct and Indirect Effects of Online Job Search Advice," .

Baker, S. R. and A. Fradkin (2017): "The Impact of Unemployment Insurance on Job Search: Evidence from Google Search Data," *The Review of Economics and Statistics*, 99, 756–768.

Banfi, S. and B. Villena-Roldan (2019): "Do High-Wage Jobs Attract More Applicants? Directed Search Evidence from the Online Labor Market," *Journal of Labor Economics*, 37, 715–746.

Behaghel, L., S. Dromundo Mokrani, M. Gurgand, Y. Hazard, and T. Zuber (2022): "Encouraging and Directing Job Search: Direct and Spillover Effects in a Large Scale Experiment," .

Belot, M., P. Kircher, and P. Muller (2018): "Providing Advice to Jobseekers at Low Cost: An Experimental Study on Online Advice," *The Review of Economic Studies*, 86, 1411–1447.

——— (2022): "Do the Long-Term Unemployed Benefit from Automated Occupational Advice During Online Job Search?" .

Bhuller, Ferraro, A. Kostol, and T. Vigtel (2023): "The Internet, Search Frictions and Aggregate Unemployment," .

Bied, G., P. Caillou, B. Crépon, C. Gaillac, E. Pérennes, and M. Sebag (2023): "Designing labor market recommender systems: the importance of job seeker preferences and competition," Mimeo.

Brown, J. and D. A. Matsa (2020): "Locked in by Leverage: Job Search During the Housing Crisis," *Journal of Financial Economics*, 136, 623 – 648.

Faberman, R. J. and M. Kudlyak (2016): "The Intensity of Job Search and Search Duration," Working Paper Series 2016-13, Federal Reserve Bank of San Francisco.

——— (2019): "The Intensity of Job Search and Search Duration," *American Economic Journal: Macroeconomics*, 11, 327–57.

Faberman, R. J., A. I. Mueller, A. Sahin, and G. Topa (2017): "Job search behavior among the employed and non-employed," *Working Paper 23731, National Bureau of Economic Research*.

Fredrickson, B. (2017): "Implicit: Fast Python Collaborative Filtering for Implicit Datasets," .

Hensvik, L., T. Le Barbanchon, and R. Rathelot (2020): "Mitigating the Work-Safety Trade-off," Working Paper DP14611, CEPR.

——— (2021): "Job search during the COVID-19 crisis," *Journal of Public Economics*, 194, 104349.

Hoffman, M., L. B. Kahn, and D. Li (2017): "Discretion in Hiring*," *The Quarterly Journal of Economics*, 133, 765–800.

Horton, J. J. (2017): "The Effects of Algorithmic Labor Market Recommendations: Evidence from a Field Experiment," *Journal of Labor Economics*, 35, 345–385.

Hu, Y., Y. Koren, and C. Volinsky (2008): "Collaborative filtering for implicit feedback datasets," in *2008 Eighth IEEE International Conference on Data Mining*, Ieee, 263–272.

Kircher, P. (2022): "Job Search in the 21St Century," *Journal of the European Economic Association*, jvac057.

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

Kudlyak, M., D. Lkhagvasuren, and R. Sysuyev (2020): "Systematic Job Search: New Evidence from Individual Job Application Data," *Journal of labor Economics*.

Marinescu, I. (2017): "The general equilibrium impacts of unemployment insurance: Evidence from a large online job board," *Journal of Public Economics*, 150, 14–29.

Marinescu, I. and R. Rathelot (2018): "Mismatch Unemployment and the Geography of Job Search," *American Economic Journal: Macroeconomics*, 10, 42–70.

Marinescu, I. and D. Skandalis (2020): "Unemployment Insurance and Job Search Behavior," .

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.

Milgrom, P. R. and S. Tadelis (2018): *How Artificial Intelligence and Machine Learning Can Impact Market Design*, 567–585, in Agrawal et al. (2019).

Naidan, B., L. Boytsov, Y. Malkov, and D. Novak (2019): "Non-Metric Space Library Manual," .

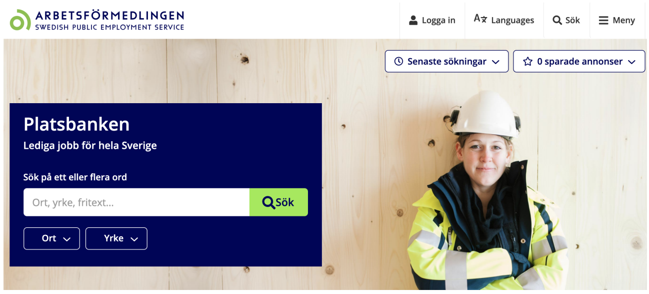
Pallais, A. (2014): "Inefficient Hiring in Entry-Level Labor Markets," *American Economic Review*, 104, 3565–3599.

Takács, G., I. Pilászy, and D. Tikk (2011): "Applications of the conjugate gradient method for implicit feedback collaborative filtering," in *RecSys '11*.

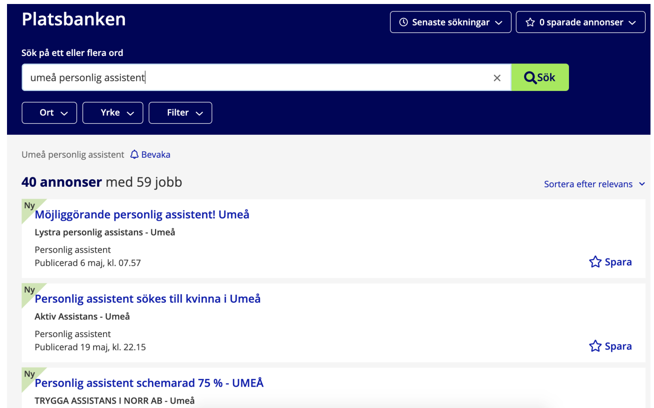
Figures

Figure 1: Screenshots from *Platsbanken.se*

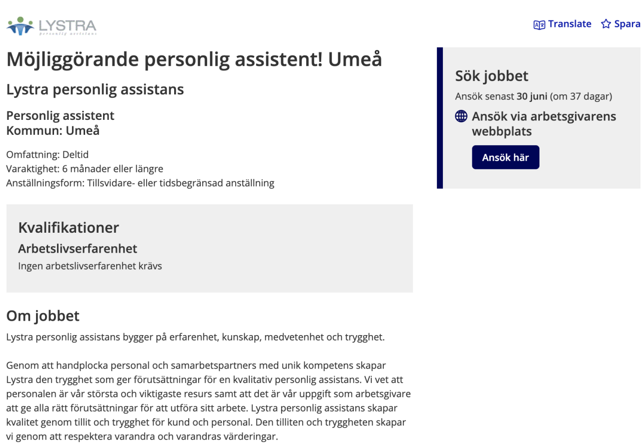
(a) Welcome page



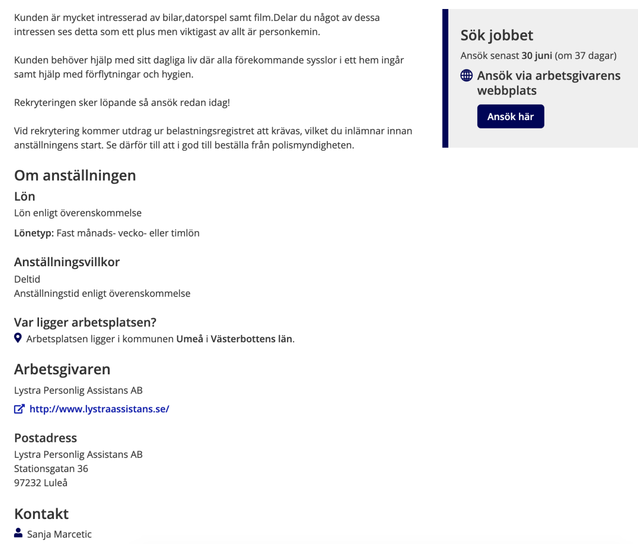
(b) Search Results



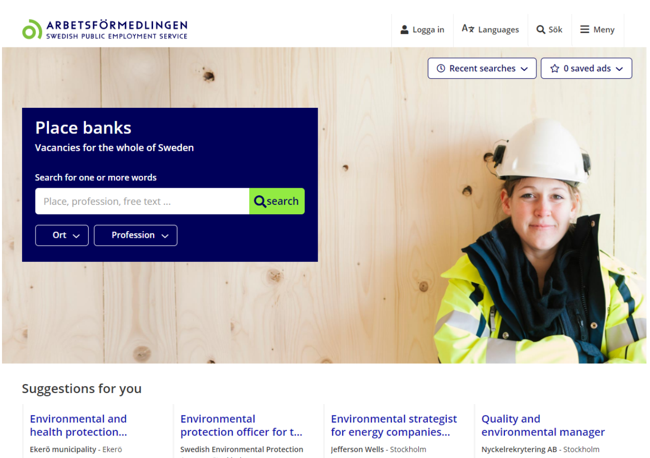
(c) Vacancy page



(d) Vacancy page (ctd)



(e) Recommendations on welcome page



(f) Recommendations on vacancy page

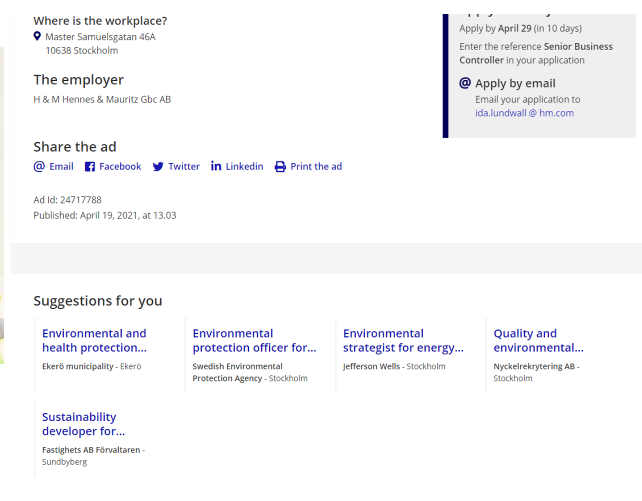
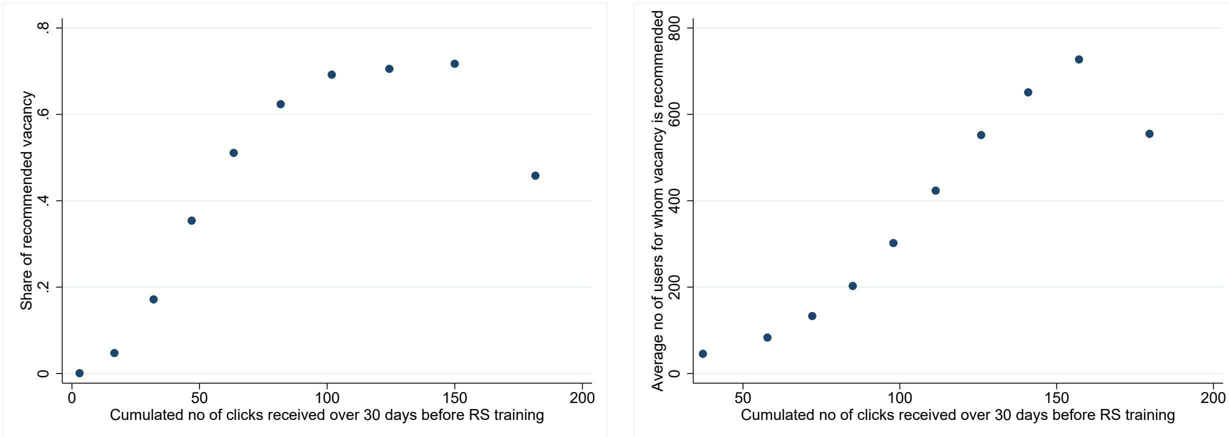


Figure 2: Popularity of recommended jobs

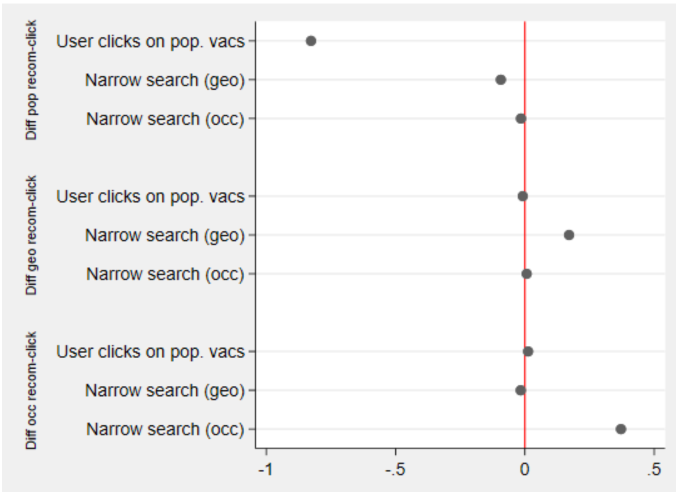


(a) Extensive Margin

(b) Intensive Margin

Note: This figure plots probability to be recommended and # times recommended for vacancies, by popularity group. Popularity is defined as the cumulated # clicks received during the 30 days before the recommender system is trained. These metrics are computed on all the vacancies posted in the first week of each month of the experimental period. Vacancies that received more than 200 clicks in the 30-days window are not considered since they would be excluded from the recommendation list.

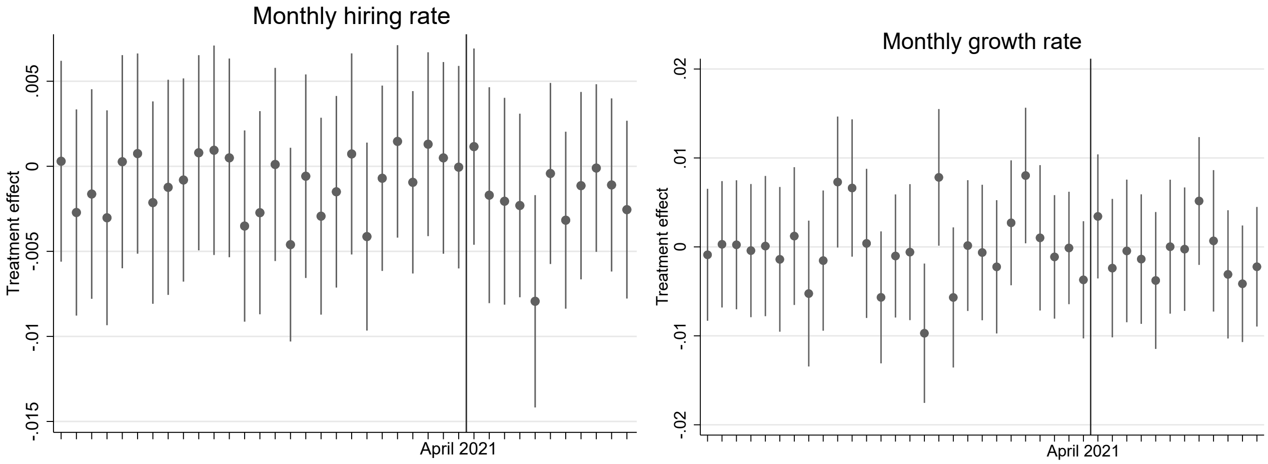
Figure 3: Recommended vacancies vs. clicked vacancies among control users, by worker characteristics



Sample: registered control users.

Note: This figure plots the coefficients of workers' search types in three different regression. We consider as search types: (i) whether the user clicks on popular vacancies (i.e that receive above median number of clicks during their first 30 days of publication), (ii) whether her geographical search is narrow (i.e. the average geo distance between clicked jobs and reference job is below median), and (iii) whether her occupational search is narrow (i.e. the average occupational transition-based distance between clicked jobs and reference job is below median). The first regression in the upper panel has as dependent variable the difference between the popularity of recommended jobs and that of clicked jobs, normalized by the standard deviation of the popularity among clicked jobs. The second regression in the middle panel considers the difference between the geographical breadth of recommended jobs and of clicked jobs, normalized by the standard deviation of breadth among clicked jobs. The third regression in the lower panel considers the difference between the occupational breadth of recommended jobs and of clicked jobs, normalized by the standard deviation of breadth among clicked jobs.

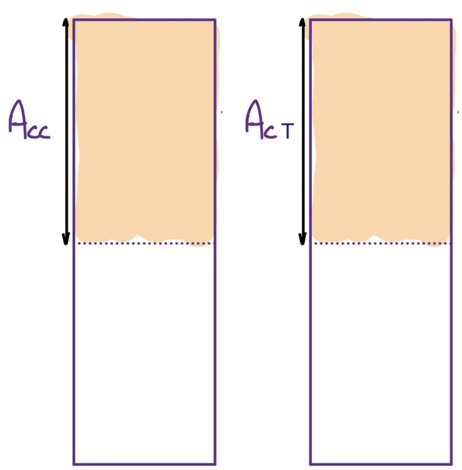
Figure 4: Effects on firms



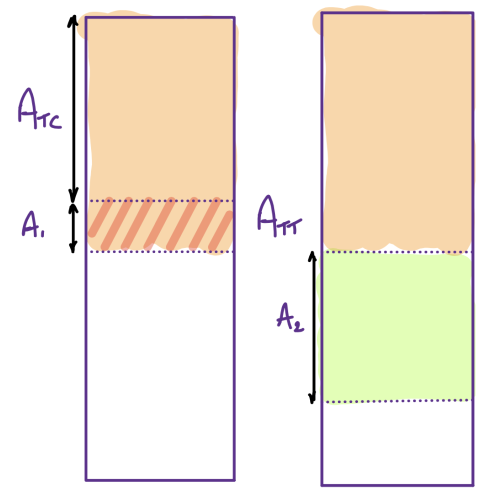
(a) Monthly Hiring Rate (b) Monthly Employment Growth Rate

Note: This figure plots the coefficient of the firm-level share of treated vacancies in a regression of firms' monthly hiring rate (panel a) and growth rate (panel b). The vertical line indicates the month when the experiment starts: April 2021.
Sample: firms with at least one experimental vacancy.

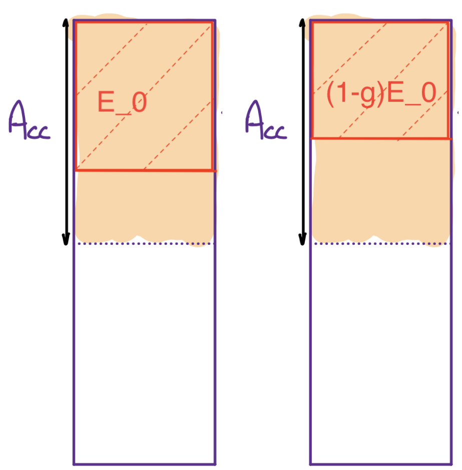
Figure 5: Decomposition of expected recommendation effects



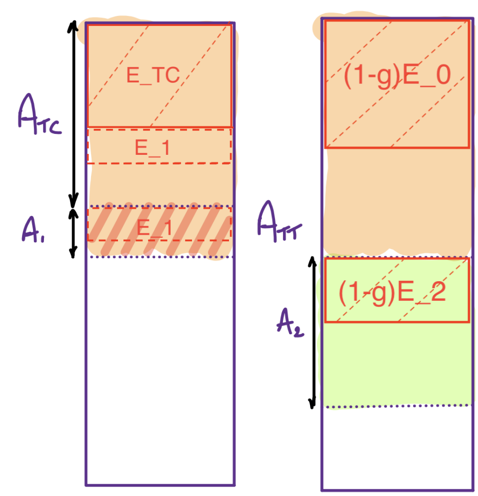
(a) Control users: Applications on control jobs (left) and on treated jobs (right)



(b) Treated users: Applications on control jobs (left) and on treated jobs (right)



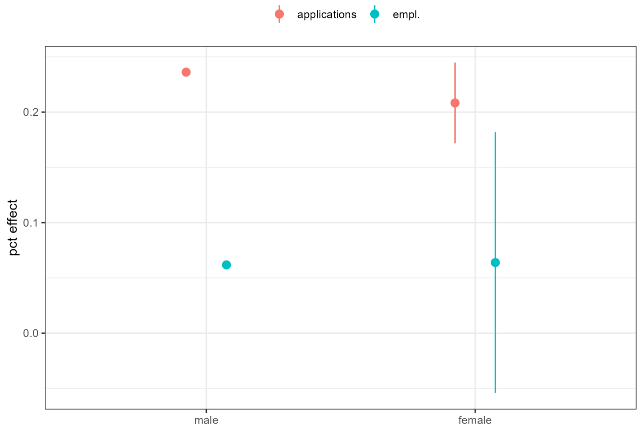
(c) Control users: Employment in control jobs (left) and in treated jobs (right)



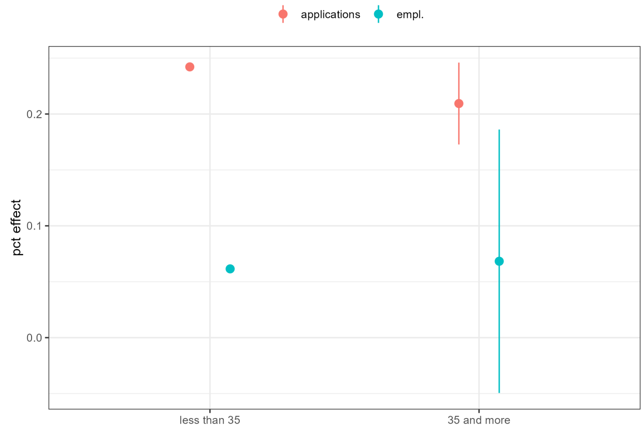
(d) Treated users: Employment in control jobs (left) and in treated jobs (right)

Note: This figure plots expected effects at the user X vacancy pair level. In each panel, the left-hand bar corresponds to control jobs, the right-hand bar to treated jobs. Both upper panels show applications, lower panels superimpose employment in solid red rectangles.

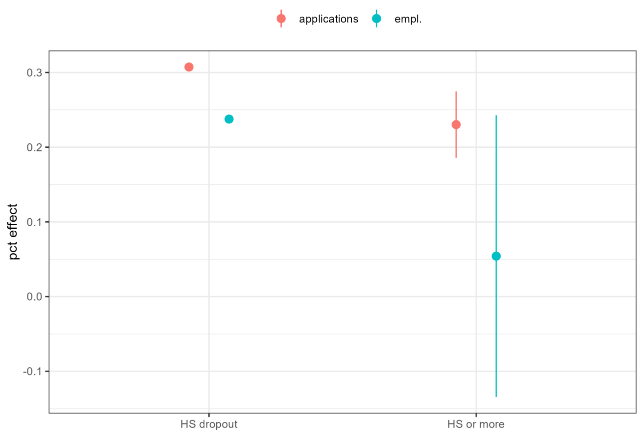
Figure 6: Heterogeneous effects by workers' characteristics



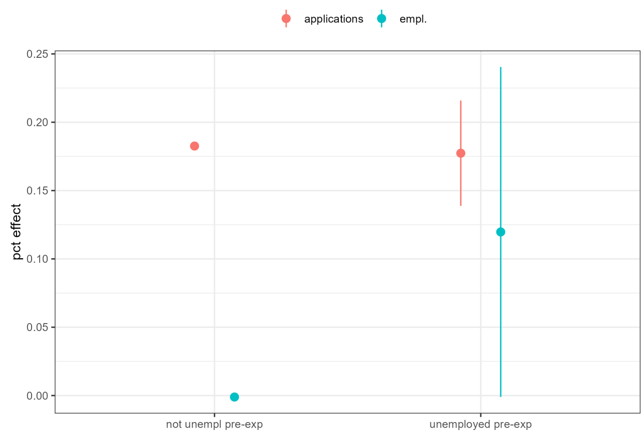
(a) Men vs Women



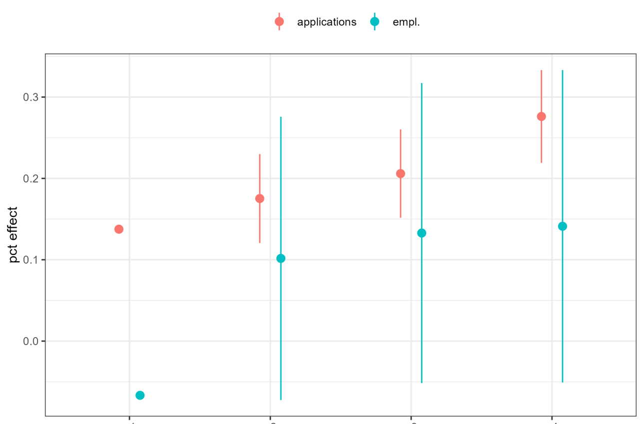
(b) Age



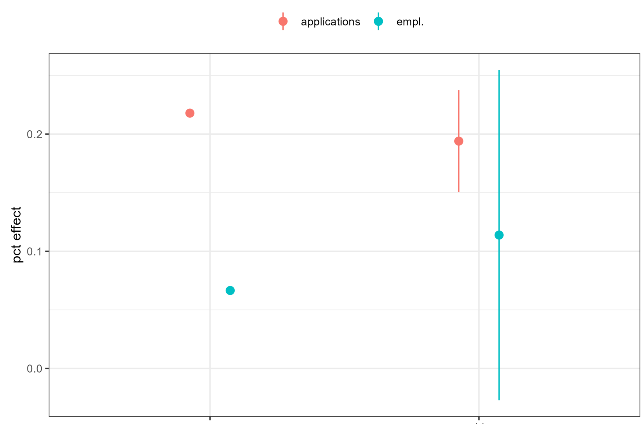
(c) Education



(d) Employed vs Unemployed



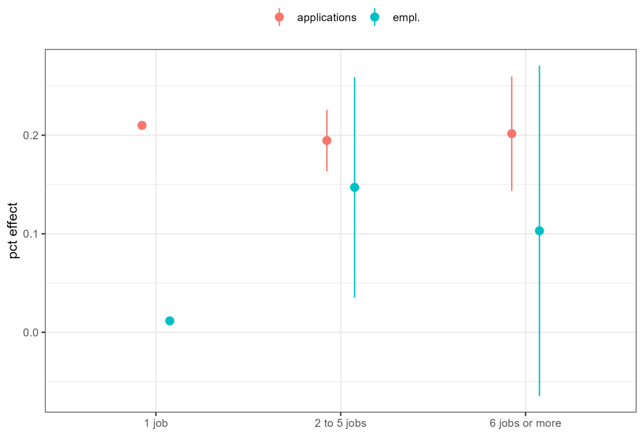
(e) Pre-exp geog. breadth of search



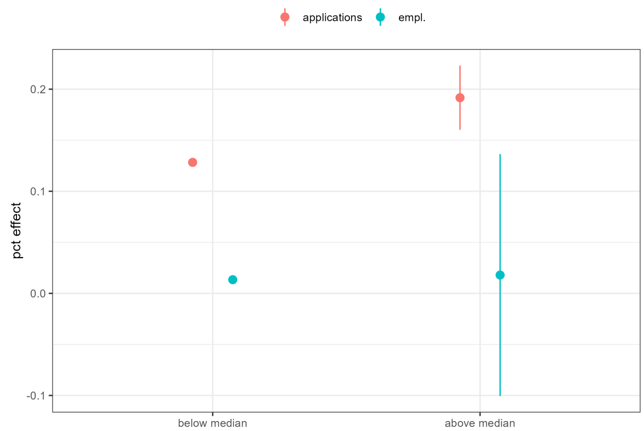
(f) Pre-exp occ. breadth of search

Note: This figure plots the pair-level treatment effects by groups of workers: men vs women in panel 6a, age in panel 6b, education in panel 6c, unemployment status in panel6d, and geographical and occupational breadth of pre-experimental search in panel 6e and 6f resp. Treatment effects for applications (in red), employment (in blue) are estimated in model (7), and represent the difference between the average outcome for treated workers on treated jobs and the average outcome for control workers on control jobs. Vertical lines represent 95% confidence interval. Standard errors clustered at the worker level and at the vacancy level.
Sample: pairs of workers-jobs such that jobs have been recommended to workers.

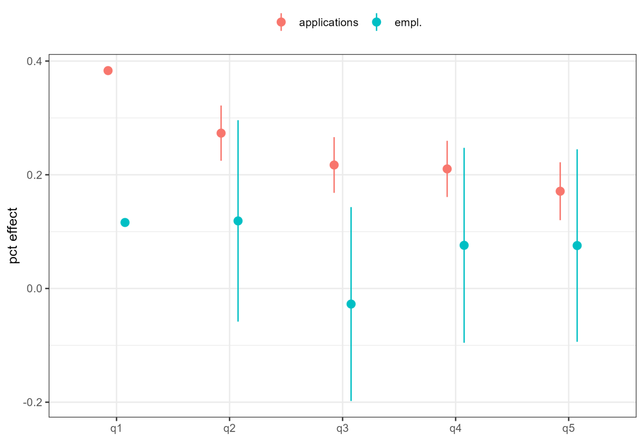
Figure 7: Heterogeneous effects by recommended vacancy characteristics



(a) No. jobs offered on ad



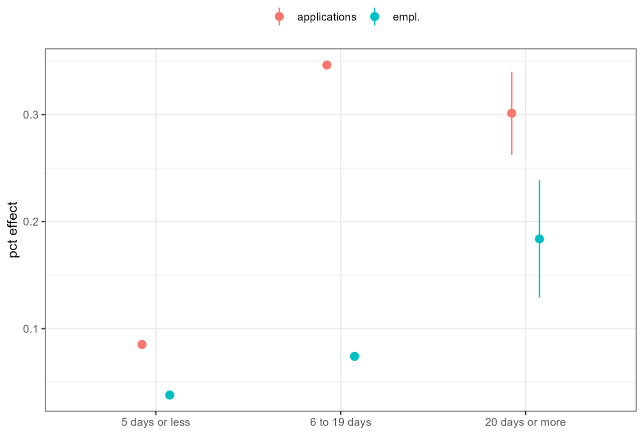
(b) No. users recommended to



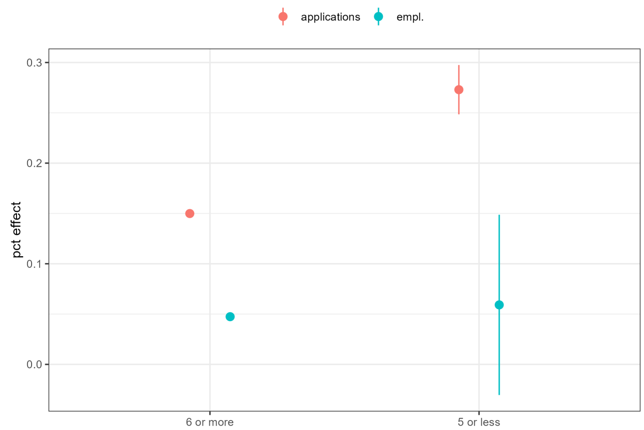
(c) Vacancy popularity

Note: This figure plots the pair-level treatment effects by groups of vacancies: # jobs offered on ad in panel 7a, total # workers vacancy is recommended to in panel 7b, and vacancy popularity in panel 7c. Treatment effects for applications (in red), employment (in blue) are estimated in model (7), and represent the difference between the average outcome for treated workers on treated jobs and the average outcome for control workers on control jobs. Vertical lines represent 95% confidence interval. Standard errors clustered at the worker level and at the vacancy level.
Sample: pairs of workers-jobs such that jobs have been recommended to workers.

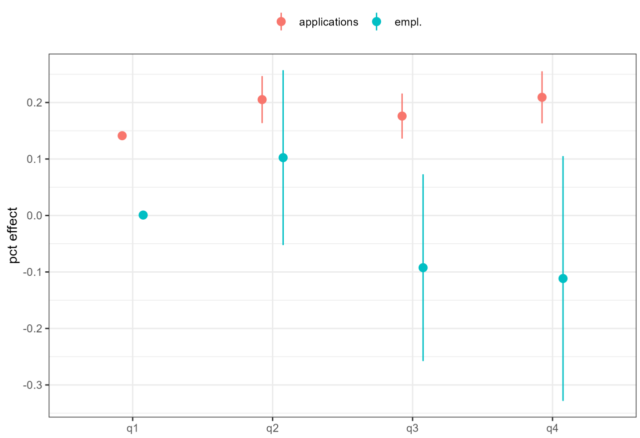
Figure 8: Heterogeneous effects by worker - recommended vacancy characteristics



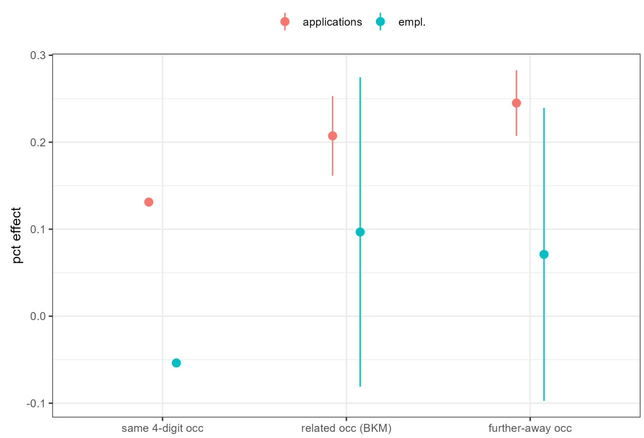
(a) Time since vacancy publication



(b) Vacancy rank within user's recom. set



(c) Distance between recommendation and users reference municipality



(d) Distance between recommendation and users reference occupation

Note: This figure plots the pair-level treatment effects by pair-level characteristics: time since vacancy first publication in panel 8a, vacancy rank within workers' recommendation set in panel 8b, and geographical and occupational distance between recommended vacancy and workers' reference job in panels 8c and 8d resp. Treatment effects for applications (in red), employment (in blue) are estimated in model (7), and represent the difference between the average outcome for treated workers on treated jobs and the average outcome for control workers on control jobs. Vertical lines represent 95% confidence interval. Standard errors clustered at the worker level and at the vacancy level.

Sample: pairs of workers-jobs such that jobs have been recommended to workers.

Tables

Table 1: Recommended vacancies vs. clicked vacancies of control users: occupation, location and popularity

	(1)	(2)	(3)	(4)	(5)
	Spontaneous activity			Recommendations	
	clicks	applications	employment	all	top 5 rank
Geographical Distance (km)	50.525	47.093	23.496	46.748	43.971
Number of same first digits as workers reference occupation code (SSYK)					
1 digit	0.322	0.353	0.361	0.293	0.308
2 digits	0.199	0.231	0.237	0.168	0.182
3 digits	0.152	0.182	0.154	0.117	0.131
4 digits	0.110	0.133	0.111	0.077	0.088
Occupational distance (classification-based)	0.804	0.775	0.784	0.836	0.823
Similar occ. (BKM)	0.149	0.180	0.213	0.130	0.140
Occupation distance (transition-based)	0.942	0.928	0.922	0.958	0.953
Popularity	12.407	13.054	12.320	9.957	11.395
Quantile of popularity	3.540	3.597	3.579	2.984	3.128
Days since publication	12.487	12.573	9.904	18.480	19.699
Obs	16,310,148	3,403,419	286,290	35,598,875	25,909,906
Individuals (knr)	247,477	199,781	50,130	247,477	247,035

Sample: control registered users during the experimental period (April 2021 - March 2022).
Note: this table reports the characteristics of vacancies that control users click, apply for, and of vacancies posted by firms hiring control users, in Columns (1) to (3) resp. In Column (4), we report the average charac. of vacancies in the recommendation set of the same control users, the day when they clicked on vacancies in Column (1). Column (5) further restricts to top ranked recommendations.
Geographical distance is between the vacancy location and the user residence. The next rows report whether the vacancy occupation code and the user stated reference occupation code share the same first digit, the same two first digits, etc. Classification-based occupational distance is computed as one minus the number of same first digits of vacancy occupation code and users reference occupation.
We consider another measure of occupational distance based on observed occupational transitions. Denote τ_{od} the share of transition to occupation d from occupation o. The distance between o and d is $1 - \tau_{od}$.
We measure vacancy popularity as the number of daily clicks received by the vacancy from the control devices during the first 30 days after publication.

Table 2: Balancing for Registered Active User

	(1) Treated	(2) Control	(3) diff	(4) pval
Women	0.493	0.493	-0.000	0.987
	0.500	0.500	0.002	.
Swedish	0.553	0.555	-0.002	0.313
	0.497	0.497	0.002	.
High-school dropouts	0.245	0.246	-0.001	0.723
	0.430	0.430	0.002	.
High-school diploma	0.420	0.423	-0.002	0.252
	0.494	0.494	0.002	.
Post-secondary educ.	0.335	0.332	0.003	0.128
	0.472	0.471	0.002	.
Stockholm resid.	0.097	0.095	0.002	0.191
	0.296	0.294	0.001	.
Before randomization month				
Monthly earnings	1035.350	1027.866	7.484	0.111
	1162.411	1162.130	4.694	.
Employment (extensive)	0.768	0.767	0.001	0.513
	0.422	0.423	0.002	.
Employment (intensive)	0.436	0.433	0.003	0.063
	0.362	0.362	0.001	.
Monthly wages	2088.273	2081.858	6.416	0.246
	1198.052	1201.271	5.531	.
During the 30 days before randomization				
No of days with clicks	3.064	3.039	0.025	0.087
	3.631	3.605	0.015	.
No of clicks	14.163	14.121	0.042	0.574
	18.537	18.477	0.075	.
No of diff. vacancies clicked	12.748	12.719	0.029	0.668
	16.533	16.526	0.067	.
No of days with apps	1.413	1.408	0.005	0.540
	2.076	2.057	0.008	.
No of apps	4.150	4.156	-0.007	0.836
	8.240	7.791	0.032	.
Observations	1.23e+05	1.22e+05	.	.

Sample: *Platsbanken.se* users randomized into treatment, registered unemployed at the time of randomization, and active after randomization.

Notes: Columns (1) and (2) report sample averages for treated and control groups respectively. Below the means are standard deviations. Column (3) report the difference in means across treated and control group. Below the differences are the standard errors. Column (4) reports the p-value for zero difference.

Monthly earnings is the average monthly earnings (incl. 0s) between January 2019 and the randomization month (excl.). Employment (intensive) is the share of months with strictly positive earnings over the same period. Monthly wages is the average monthly wage (excl. months with zero earnings) over the same period.

Clicks and applications are for vacancies posted on *Platsbanken.se*, and cumulated over the month before randomization into treatment.

Table 3: Effect on daily clicks and daily applications per user

	(1)	(2)	(3)	(4)
	All	Recommended jobs		
	jobs	Yes	Yes but not shown	No
		Panel A: Clicks		
User is treated	0.00451	0.0463***	-0.00111***	-0.0407***
	(0.00526)	(0.000417)	(0.000325)	(0.00514)
Control mean	3.347	0.105	0.106	3.136
Outcome sd	3.255	0.363	0.366	3.128
pct impact	0.135	44.08	-1.048	-1.297
		Panel B: Applications		
User is treated	-0.00368*	0.00438***	-0.000268**	-0.00779***
	(0.00214)	(0.000132)	(0.000113)	(0.00201)
Control mean	0.400	0.0143	0.0144	0.371
Outcome sd	1.096	0.131	0.131	1.038
pct impact	-0.921	30.66	-1.864	-2.101
		Panel C: Within-day conversion rates (in %)		
Control Users	11.9	13.6	13.5	11.8
	(0.050)	(0.091)	(0.093)	(0.047)
Treated Users	11.8	12.3	13.4	11.7
	(0.050)	(0.083)	(0.091)	(0.048)
Observations	14,605,215	14,605,215	14,605,215	14,605,215
no of users	1.720e+06	1.720e+06	1.720e+06	1.720e+06

Sample: active users (i.e., who click at least once on day d)

Note: this table reports the treatment effects from user-day level regressions. For each user, we consider all post-randomization days when she clicks on at least one vacancy. In Panel A, we report treatment effects on the daily number of clicks per user (column 1), restricting to clicks on treated vacancies in users recommendation set in Column (2). Column (3) restricts to clicks on control vacancies in users recommendation set, which are not shown. Column (4) considers clicks on the complement sample of vacancies (user-specific complement). We report treatment effects both in absolute value and in percentage (as the ratio of absolute effect on the corresponding average among control users). In Panel B, we report treatment effects on daily applications. In Panel C, we compute conversion rates from clicks to applications as implied by panels A and B.

Robust standard errors clustered at user level in parenthesis. They are computed with delta method in Panel C and E. The conversion rate is obtained as the ratio of the application estimate ($\hat{A}$) and of the click estimate ($\hat{C}$). Let us denote $se(C)$ (resp. $se(A)$) the standard errors of $\hat{C}$ (resp. $\hat{A}$). Then using the delta method, we obtain the standard errors of the ratio ($\hat{R} = \hat{A}/\hat{R}$) as: $se(R) = \left(se(A)^2 / \left(\hat{C}\right)^2 + se(C)^2 \left(\hat{A}\right)^2 / \left(\hat{C}\right)^4\right)^{1/2}$.

Table 4: Effect on reemployment outcomes of job seekers

	(1) Received earnings	(2) Monthly earnings	(3) Empl. duration
	Panel A: Baseline Specification		
User is treated	0.00389**	5.234	0.00265*
	(0.00197)	(4.224)	(0.00158)
pct impact	0.641	0.630	0.681
	Panel B: with controls		
User is treated	0.00378**	3.720	0.00225
	(0.00178)	(3.782)	(0.00143)
User controls	X	X	X
	Panel C: Double-Debiased Machine-Learning estimator		
User is treated	0.00313*	2.179	0.00174
	(0.00177)	(3.733)	(0.00141)
User controls	X	X	X
Observations	245,209	245,209	245,209
Control mean	0.606	830.4	0.389

Sample: Registered users (active after randomization).
This table reports treatment effects on the fraction of months that users received any labor earnings between the first randomization month and the end of employment register dataset (column 1), on the average monthly earnings (column 2), and on the fraction of month that users are continuously employed in the same firm (column 3). After reporting the treatment effects in levels, we provide percentage impact from control means.
Panel A does not control for any covariates, while in Panel B we conclude as controls: average employment, monthly earnings, no. of clicks, no. of applications, no. of different vacancies clicked and no. of days with applications, all considered before the experiment, and dummies for female, Swedish, high school dropouts, having a high school diploma, post-secondary education, Stockholm. Municipality, occupation and month of randomization fixed effects.
Robust standard errors in parenthesis.

Table 5: Effect on daily clicks and applications received per vacancy

	(1)	(2)	(3)	(4)
		users with job in recomm. set		
	all users	all	treated	control
Panel A: Daily clicks received				
Vacancy is treated	0.0953***	0.0887***	0.0875***	-0.00158*
	(0.0223)	(0.00181)	(0.00103)	(0.000833)
Control mean	8.480	0.360	0.179	0.180
Outcome sd	9.547	1.058	0.608	0.609
pct impact	1.124	24.66	48.80	-0.879
Panel B: Daily applications received				
Vacancy is treated	0.0132***	0.0115***	0.0118***	-0.000325
	(0.00360)	(0.000535)	(0.000314)	(0.000277)
Control mean	0.655	0.0757	0.0375	0.0380
Outcome sd	2.041	0.480	0.294	0.305
pct impact	2.018	15.25	31.57	-0.854
Observations	9,078,687	9,078,687	9,078,687	9,078,687
no of jobs	605114	605114	605114	605114

Sample: experimental vacancies (excl. super control markets). Vacancy x day analysis, where we include all days when vacancy is recommended at least once to an experimental user (either control or treated).

Note: This table reports treatments effects on daily clicks received per vacancy in panel A and daily applications received in Panel B. We winsorize daily clicks from each user type separately. In Column (1), we count clicks and applications from any users. From Column (2) onwards, we count clicks from users for whom the vacancy belongs to their daily recommendation set. In Columns (3) and (4), we further split the users population between treated and control users. We report treatment effects in levels, and in percentage (level effects over the control mean). Robust standard errors clustered at vacancy-level in parenthesis.

Table 6: Super-treatment effect on control vacancies and their firms

	(1) Clicks Received	(2) Applications Received	(3) Log Firm Employment
Market is treated	-0.327	-0.0192	0.0355
	(0.261)	(0.0164)	(0.0226)
Observations	1,249,828	1,249,828	10437
No. firms	10437	10437	10437
R-squared	0.453	0.183	0.772
Control mean	7.217	0.441	
Outcome sd	8.082	1.073	1.860
pct impact	-4.533	-4.363	1.079

Sample: all control vacancies posted between November 2021 and March 2022 (excl. STO commuting zone), and firms that post at least one control vacancies in Column (3).

Note: This table reports effects of belonging to a treated market (as opposed to super control markets). Markets are defined by commuting zone X occupational skill groups (High vs Low). We report the market-treatment effects on daily clicks received in Column (1), and on daily applications received in Column (2). The underlying regression is at the day X vacancy level. We report the market-treatment effects on firm-level employment in Column (3) (where observations are weighted by the inverse of vacancies posted by firms times the number of days they appear in the sample).The regression models include fixed effects for randomization market pair. Robust standard errors clustered at market-level in parenthesis.

Table 7: Worker × recommended job pair-level analysis

	(1) Clicks	(2) Applications	(3) Employment	(4) Emp. duration	(5) Earnings
	Panel A: Mean outcomes by users X vacancy treatment groups				
Treated user, Treated vacancy (α_{11})	0.0479***	0.0123***	0.000490***	0.00227***	4.331***
	(5.76e-05)	(2.93e-05)	(5.85e-06)	(3.26e-05)	(0.0727)
Treated user, Control vacancy (α_{10})	0.0340***	0.00993***	0.000450***	0.00213***	4.083***
	(4.81e-05)	(2.62e-05)	(5.57e-06)	(3.15e-05)	(0.0692)
Control user, Treated vacancy (α_{01})	0.0345***	0.0102***	0.000457***	0.00216***	4.117***
	(4.89e-05)	(2.69e-05)	(5.70e-06)	(3.22e-05)	(0.0697)
Control user, Control vacancy (α_{00})	0.0344***	0.0102***	0.000464***	0.00220***	4.288***
	(4.86e-05)	(2.66e-05)	(5.67e-06)	(3.20e-05)	(0.0711)
	Panel B: Pair-level effect (Treated - control pairs)				
Pair-level effect ($\alpha_{11} - \alpha_{00}$)	0.0134***	0.00211***	2.56e-05***	6.65e-05	0.0433
	(7.30e-05)	(3.87e-05)	(8.02e-06)	(4.49e-05)	(0.0997)
pct impact	38.96	20.70	5.516	3.021	1.011
Lower CI bound	38.54	19.95	2.128	-0.973	-3.548
Upper CI bound	39.37	21.44	8.904	7.015	5.569
Observations	59,243,200	59,243,200	59,243,200	59,243,200	59,243,200

Sample: registered users included in the experimental sample (excl. super control markets) and their list of recommended vacancies. User x vacancy pair-level analysis.

Note: this Table reports in Panel A pair-level means on search activity (Columns 1 and 2) and on matching outcomes (Columns 3 onwards). They are the αs coefficients from Regression (7): $Y_{ij} = \alpha_{00}\mathbb{1}(T^u_{ij} = 0)\mathbb{1}(T^v_{ij} = 0) + \alpha_{01}\mathbb{1}(T^u_{ij} = 0)\mathbb{1}(T^v_{ij} = 1) + \alpha_{10}\mathbb{1}(T^u_{ij} = 1)\mathbb{1}(T^v_{ij} = 0) + \alpha_{11}\mathbb{1}(T^u_{ij} = 1)\mathbb{1}(T^v_{ij} = 1) + \varepsilon_{ij}$. Robust standard errors. In Panel B, we report the pair-level effect, ie the difference between α_{11} and α_{00}, in levels and in percent impact.

All clicks and applications on recommended jobs after the first day it appears in the recommendation lists are considered. Employment is in firms that posted vacancy recommended to that user and for which users applied for. Emp. duration is the number of months of employment between randomisation and June 2022. Earnings are total earnings.

Table 8: Impact of treatment on workers' probability of employment and congestion effects: pair-level analysis

Component	Estimate	p-value	Comment
Effect T jobs	0.035 (0.009)	0.000	$(E_{TT} - E_{CT})/2E_0$
Effect C jobs	-0.015 (0.008)	0.075	$(E_{TC} - E_{CC})/2E_0$
Net effect	0.02 (0.012)	0.103	Sum of row 1 and 2
Congestion g	0.014 (0.017)	0.401	$1 - E_{CT}/E_0$

Authors' calculations. Section 6 details the definition of theoretical quantities E_{TT}, E_{TC}, E_{CT}, E_{CC}, and g and Section D explains how they are obtained from estimates showed in Table 7 and how the impact of treating workers on treated and control jobs are computed from these quantities (column Comment summarises the procedure).
Standard errors are heteroskedasticity robust.
The stars indicate the significance of the estimates based on a t-test.

Online Appendix

Job Recommender System

Lena Hensvik, Thomas Le Barbanchon, Roland Rathelot

In Appendix A, we provide an ex-ante analysis of job recommender system properties. We provide details on the design of the market-level randomization in Appendix B. We document balancing tests of various experimental samples in Appendix C. Appendix E includes extra figures, while Appendix F includes extra tables.

A Ex-ante properties of job recommender system: Mean Average Precision

We average over all users u the following score MAP(u). We define first the ranked list of recommendations $(rec(1), ..., rec(k))$ of size k where $rec(1)$ is the most relevant vacancy according to the recommender system. $C(u)$ is the set of clicked vacancies the day when the recommendations would have been shown. Then the precision criteria MAP(u) writes:

$$MAP = \frac{1}{\min(k, Card(C(u)))} \sum_{i=1}^{k} \mathbb{1}\left[rec(i) \in C(u)\right] \frac{\sum_{j=1}^{i} \mathbb{1}\left[rec(j) \in C(u)\right]}{i} \tag{9}$$

In a nutshell, this looks at how many of these recommendations users actually clicked during the potential exposure day. In Figure A1, we compute the MAP varying the number of recommendations to each user. First, when only one job is recommended, 3% of users click on it. Then, when we recommend two vacancies to each user or more, the score interpretation is more involved. The score features an implicit weighting scheme that gives more importance to the first recommendations in the list. Broadly speaking, if a recommendation is clicked, it increases the score.

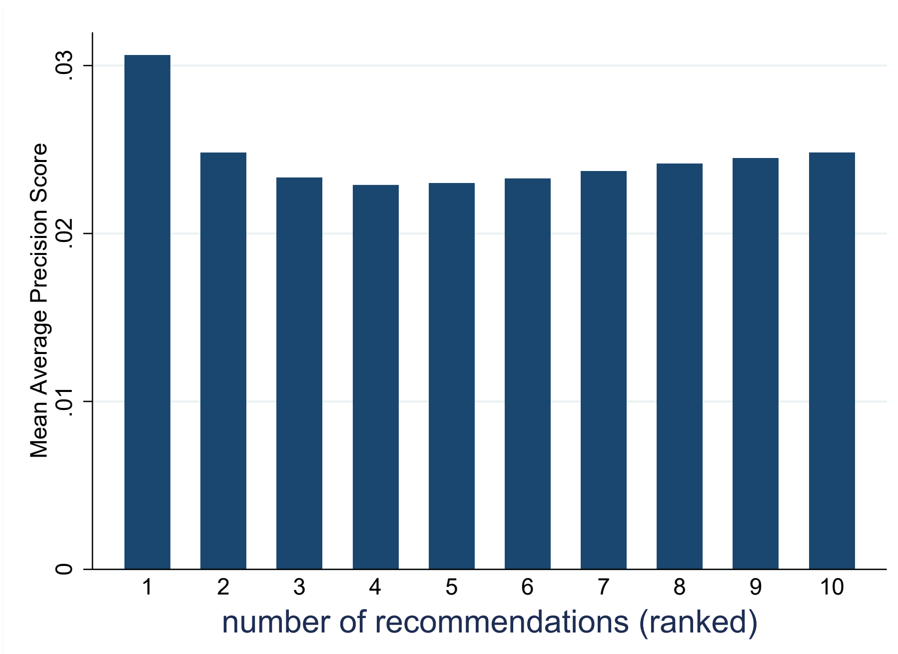


Figure A1: Mean Average Precision

Note:

B Market-level randomization

In November 2021, we introduced a new market-level randomization plan, by assigning Sweden's local labor markets to either a Super Treatment (ST) or Super Control (SC) group. The markets are defined at the *commuting zone* $\times$ *skill* level, where skill is a categorical variable equal to *high* if the vacancy's occupation belongs to one of the first three SSYK2012[16] major groups, namely "managers", "occupations requiring advanced level of higher education" or "occupations requiring higher education qualifications or equivalent", and equal to *low* otherwise. Thus, each of the 70 Swedish commuting zones is interacted with the skill variable, resulting in 140 local labor markets.

To carry out the randomization plan, we use a constrained K-means algorithm, generating constrained $CZ \times skill$ pairs based on three outcome statistics computed over the months of April and May 2021: average conditional daily number of clicks per vacancy, average conditional daily number of applications per vacancy and average daily number of vacancies with at least one click, by local labor market. We then randomize within unit-pair and assign each element to either Super Treatment or Super Control. Notably, Stockholm's commuting zone is excluded from this process and is always assigned to the Super Treatment group.

Figure B1 plots the results of the market-level randomization.

Table B1 shows the balancing of vacancies' characteristics across the two groups, considering all vacancies between December 2021 and March 2022. Table B3 shows the same balancing but considering only control vacancies.

While the market-level randomization allows to directly obtain the super treatment status of each vacancy in the experimental sample, for devices it is not as clear cut. To separate users into a super treatment and a super control group, we obtain their reference local labor market, defined as the modal labor market of the vacancies they click on in the 30-days window before they enter the experiment. Thus, users do not have a Super Treatment or Super Control status, but a reference ST or SC one. Figure B2 shows the share of vacancies in the user recommendation set from the Super Treatment group, by users' reference super treatment status. The asymmetry between the two subfigures is due to Stockholm's local labor markets being assigned by default to the Super Treatment group: considering vacancy recommendations from outside the users' reference local labor market, it is more likely that they are from ST markets. Indeed, taking out recommendations from Stockholm's commuting zone the two histograms become symmetric.

[16] https://www.scb.se/contentassets/0c0089cc085a45d49c1dc83923ad933a/structur_english_ssyk2012_1_4.xlsx

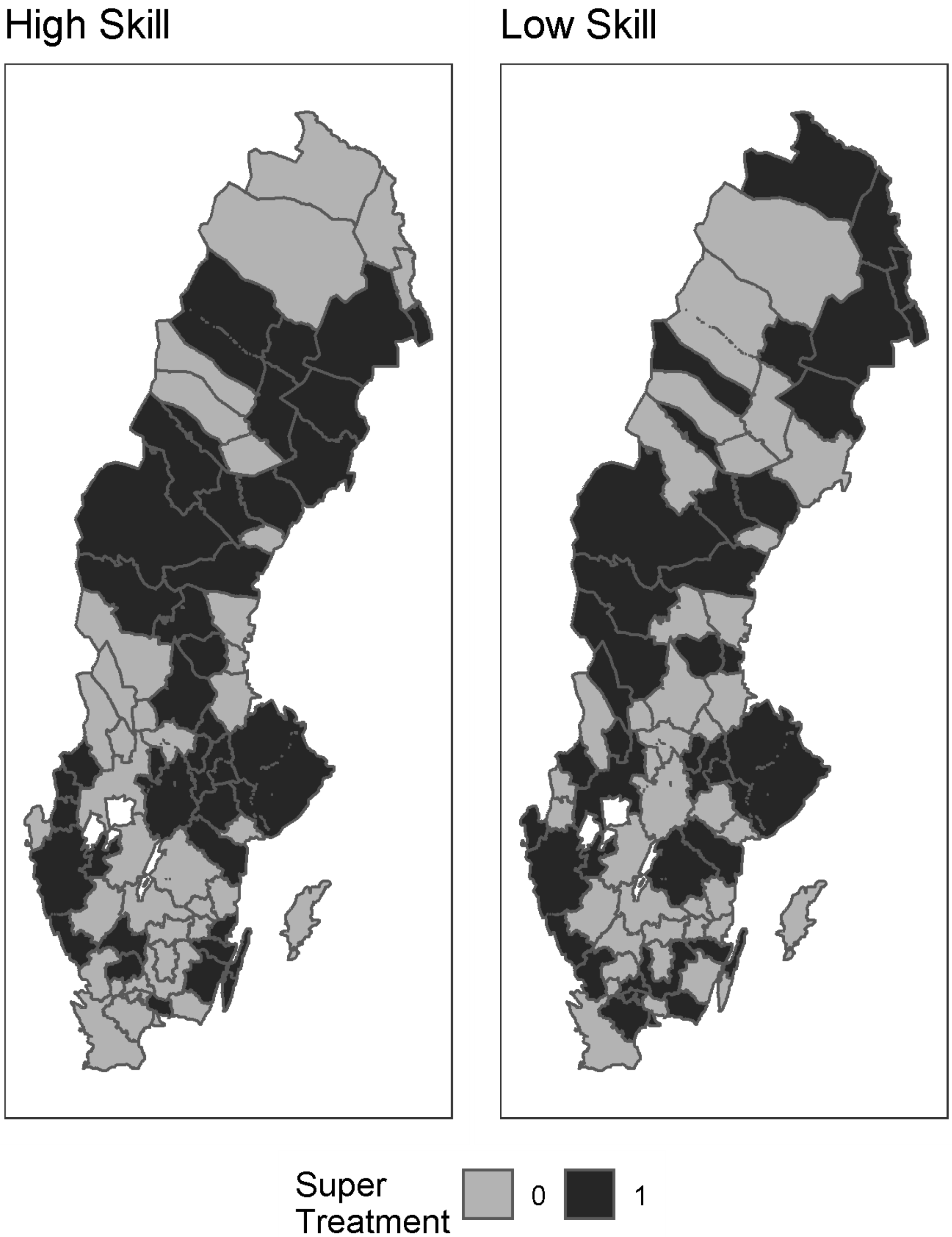


Figure B1: Distribution of Super Treatment and Super Control areas

Note: This figure shows the Super Treatment status of each local labor market (defined at the $CZ \times skill$ level). The subfigure on the left shows the local labor markets for high skill occupations, while the one on the right plots the low skilled ones. Dark colored areas are randomized into Super Treatment, while lighter ones into Super Control. The randomization is carried out through a constrained K-means algorithm based on three outcome statistics: average conditional daily number of clicks per vacancy, average conditional daily number of applications per vacancy and average daily number of vacancies with at least one click, by market

Table B1: Balancing for local labor markets

	ST	SC	diff	pval
Days out before rand.	3.758	3.727	.032	.785
Total no of clicks before rand.	67.569	66.529	1.040	.419
Average no of daily clicks before rand.	37.191	36.895	.296	.819
Total no of app. before rand.	3.397	3.401	-.005	.979
Average no of daily app. before rand.	2.890	2.930	-.040	.709
Open-ended contract	0.655	0.648	.007	.586
Regular employment	0.897	0.898	-.001	.959
Fixed salary	0.934	0.937	-.003	.593
Full-time	0.762	0.761	.002	.902
Requires experience	0.689	0.716	-.026	.117
Nurse occ.	0.072	0.066	.006	.483
Teacher occ.	0.060	0.062	-.002	.737
ICT architect occ.	0.020	0.020	.001	.858
Admin and support ind.	0.192	0.195	-.003	.820
Health and social ind.	0.346	0.339	.007	.689
Observation	69	69		

Notes: Columns (1) and (2) report sample averages for super treatment and super control areas respectively. Below the means are standard deviations. Column (3) reports the difference in means across super treatment and super control groups. Column (4) reports the p-value for zero difference, obtained from the following OLS specification: $Y_m = \delta ST_m + \mu_p + \varepsilon_m$, where μ_p are the randomization pair fixed effects and m is the local labor market identifier. We employ robust standard errors.
Sample: all vacancies from December 2021 to March 2022.

Table B3: Balancing for local labor markets

	ST	SC	diff	pval
days out before rand.	3.652	3.631	0.021	0.888
Total no of clicks before rand.	66.611	66.009	0.602	0.610
Average no of daily clicks before rand.	37.933	37.403	0.530	0.694
Total no of applications before rand.	3.306	3.390	-0.085	0.580
Average no of daily apps before rand.	2.934	3.018	-0.084	0.495
Open-ended contract	0.652	0.643	0.009	0.435
Regular employment	0.896	0.907	-0.010	0.213
Fixed salary	0.936	0.936	-0.000	0.951
Full-time	0.757	0.763	-0.007	0.608
Requires experience	0.695	0.709	-0.014	0.444
Nurse occ.	0.077	0.070	0.007	0.502
Teachers occ.	0.059	0.062	-0.003	0.425
ICT architect occ.	0.019	0.018	0.001	0.730
Admin and support ind.	0.199	0.209	-0.010	0.463
Health social ind.	0.347	0.349	-0.002	0.920
Observation	68	68		

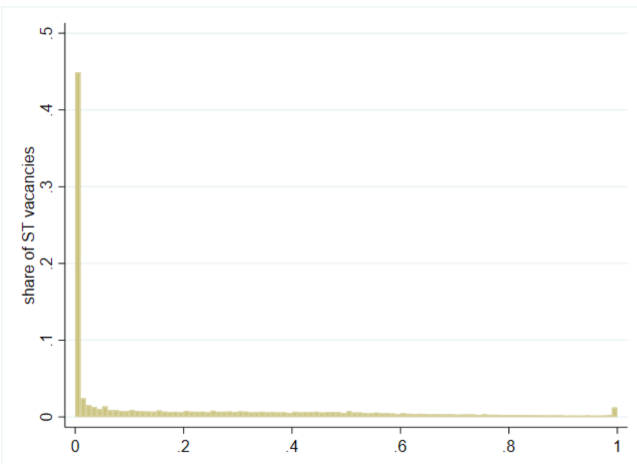
Notes: Columns (1) and (2) report sample averages for super treatment and super control areas respectively. Below the means are standard deviations. Column (3) reports the difference in means across super treatment and super control groups. Column (4) reports the p-value for zero difference, obtained from the following OLS specification: $Y_m = \delta ST_m + \mu_p + \varepsilon_m$, where μ_p are the randomization pair fixed effects and m is the local labor market identifier. We employ robust standard errors.
Sample: all control vacancies from November 2021 to March 2022.

Table B5: Balancing for local labor markets

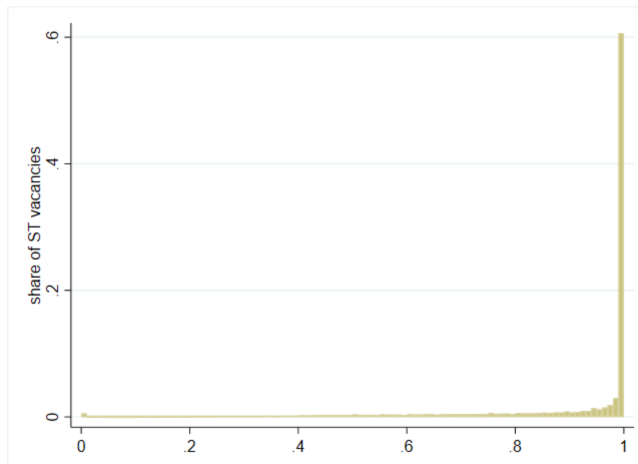
Variable Names	T1	T0	diff	pval
Days out before randomization	4.401	3.971	0.579	0.287
Total no of clicks before randomization	63.453	62.633	0.808	0.712
Average no of daily clicks before randomization	32.078	33.389	-1.743	0.382
Total no of applications before randomization	2.448	2.606	-0.269	0.427
Average no of daily applications before randomization	1.742	1.915	-0.274	0.308
Open-ended contract	0.641	0.578	0.049	0.089
Regular employment	0.896	0.888	0.004	0.814
Fixed salary	0.936	0.929	0.005	0.785
Full-time	0.736	0.729	0.005	0.878
Requires experience	0.620	0.652	-0.045	0.365
Nurse occ.	0.111	0.098	0.026	0.448
Teacher occ.	0.069	0.076	-0.007	0.520
ICT architect occ.	0.005	0.005	-0.001	0.827
Admin and support ind.	0.144	0.151	-0.012	0.737
Health and social ind.	0.438	0.478	-0.028	0.421
Observation	68.000	68.000	.	.

Notes: Columns (1) and (2) report sample averages for super treatment and super control areas respectively. Below the means are standard deviations. Column (3) reports the difference in means across super treatment and super control groups. Column (4) reports the p-value for zero difference, obtained from the following OLS specification: $Y_m = \delta ST_m + \mu_p + \varepsilon_m$, where μ_p are the randomization pair fixed effects and m is the local labor market identifier. We employ robust standard errors.

Sample: all control vacancies from November 2021 to March 2022. Observations are weighted by the inverse of the number of vacancies posted in each local labor market.



(a) Reference Super Control



(b) Reference Super Treatment

Figure B2: Share of ST vacancies in ST labor markets by users' reference labor market

Note: This figure plots the share of ST vacancies in the recommendation out of all potential recommendations of a user, by users' reference local labor market. Sample: all recommendations in the month of November 2021.

C Balancing tables

C.1 User analysis

We test for balancing between treated and control users included in the experimental analysis from April 2021 to March 2022. This corresponds to the main sample $\mathcal{S}_a$ in the main text. Note that users in $\mathcal{S}_a$ are *active*, as they click on at least one vacancy in a day when recommendations are generated for them. First, we focus on the job search history on *Platsbanken.se* before users are randomized into the recommender system. Those pre-randomization variables are available whether users are registered as unemployed or not. In Table C2, we analyze the total number of clicks/views and of applications on vacancies during the 30 days before randomization. Beyond those measures of search intensity, we analyze the average characteristics of the clicked vacancies (location, contract type, hours worked, experience requirement, occupation, industry). Then, for each user, we define a reference occupation (resp. municipality) as the modal occupation (resp. municipality) clicked over the 30 days before randomization. In Table C1, we test for balance among those categorical variables. Overall, we find that pre-randomization variables are balanced across treated and control groups.

Table C1: Balancing for users: categorical variable

	Share of p-values > 0.05
Occupation reference	0.95
municipality reference	0.97

Notes: This table reports the share of treated-control differences across all categories with p-value >0.05. Sweden has 290 municipalities. There are over 400 4-digit occupational categories in Swedish Standard Occupation Classification (SSYK).

Table C2: Balancing for users

Variable Names	T1	T0	diff	pval
Total no clicks	7.538	7.543	-0.005	0.393
	4.113	4.122	0.006	.
No active days	1.671	1.671	0.000	0.812
	0.960	0.965	0.001	.
Average no daily clicks	1.100	1.099	0.000	0.696
	0.191	0.190	0.000	.
Total no applications	3.529	3.531	-0.002	0.883
	8.820	8.832	0.013	.
Average no daily applications	2.189	2.185	0.004	0.272
	1.868	1.866	0.004	.
Average no vacancies clicked	5.317	5.327	-0.009	0.134
	4.019	4.038	0.006	.
region stockholm	0.225	0.225	0.000	0.809
	0.391	0.391	0.001	.
region vastra	0.164	0.165	-0.000	0.905
	0.351	0.352	0.001	.
Open-ended contract	0.631	0.631	-0.000	0.998
	0.267	0.268	0.000	.
Regular employment	0.896	0.895	0.000	0.229
	0.182	0.183	0.000	.
Fixed salary	0.950	0.950	0.000	0.015
	0.113	0.114	0.000	.
Full-time	0.694	0.694	0.001	0.163
	0.298	0.298	0.000	.
Requires experience	0.732	0.731	0.000	0.336
	0.254	0.254	0.000	.
Nurse occ.	0.020	0.020	0.000	0.524
	0.116	0.116	0.000	.
Teacher occ.	0.043	0.043	0.000	0.961
	0.161	0.161	0.000	.
ICT architect occ.	0.018	0.019	-0.000	0.393
	0.103	0.103	0.000	.
Admin and support ind.	0.199	0.199	0.000	0.172
	0.228	0.228	0.000	.
Health and social ind.	0.209	0.208	0.001	0.175
	0.277	0.276	0.000	.
Temporary help ind.	0.169	0.168	0.000	0.338
	0.215	0.215	0.000	.
Unique users	8.58e+05	8.55e+05	.	.

Notes: Columns (1) and (2) report sample averages for treated and control groups respectively. Below the means are standard deviations. Column (3) report the difference in means across treated and control group. Below the differences are the standard errors. Colmun (4) reports the p-value for zero difference.

C.2 Registered job seeker analysis

We perform a similar balancing analysis on the subsample of experimental users who registered at *Arbetsformedlingen* $\mathcal{S}_a$. Note that as in the previous subsection, this sample is restricted to active users. However, we differ from the daily activity definition, as the analysis is at the worker-level. We consider as active, users who click on at least one vacancy after the first day when recommendations are generated. Among the overall population of registered users, the probability that users visit *Platsbanken.se* website after their randomization is 81.6%, and strictly equal across experimental arms. We comment the balancing table for active registered user, Table 2, in the main text. Table C4 report the results of balancing tests for categorical variables.

Table C4: Balancing for active registered user: categorical variables

	Share of p-values $>$ 0.05
Occupation reference	0.94
municipality reference	0.95

Notes: This table reports the share of treated-control differences across all categories with p-value $>$0.05. Sweden has 290 municipalities. There are over 400 4-digit occupational categories in Swedish Standard Occupation Classification (SSYK). Sweden has 290 municipalities.

C.3 Vacancy analysis

Tables C5 and C6 report balancing tests on the sample of vacancies included in the experiment. They are all vacancies that appear at least once in the recommendation set of a user. We exclude from the samples vacancies that belong to super control local labor markets. See previous Section for the definition of super control local labor markets.

Table C5: Balancing for vacancies

	Treatment	Control	Difference	P-value
Days out before randomization	4.064	4.063	0.001	0.953
	5.706	5.723	0.015	.
Total no of clicks before rand.	72.252	72.341	-0.089	0.420
	42.970	42.874	0.110	.
Average no of daily clicks before rand.	37.949	38.059	-0.110	0.243
	36.711	36.749	0.094	.
Total no of applications before rand.	5.729	5.741	-0.012	0.564
	8.048	7.945	0.021	.
Average no of daily apps before rand.	3.966	3.974	-0.008	0.631
	6.420	6.386	0.016	.
region stockholm	0.291	0.288	0.003	0.026
	0.454	0.453	0.001	.
region vastra	0.174	0.176	-0.001	0.126
	0.379	0.381	0.001	.
Open-ended contract	0.682	0.681	0.001	0.567
	0.466	0.466	0.001	.
Regular employment	0.915	0.914	0.001	0.053
	0.279	0.281	0.001	.
Fixed salary	0.931	0.930	0.000	0.462
	0.253	0.254	0.001	.
Full-time	0.773	0.772	0.001	0.292
	0.419	0.420	0.001	.
Requires experience	0.761	0.761	0.000	0.721
	0.426	0.427	0.001	.
Nurse occ.	0.045	0.046	-0.000	0.579
	0.208	0.209	0.001	.
Teacher occ.	0.050	0.050	0.000	0.894
	0.217	0.217	0.001	.
ICT architect occ.	0.050	0.050	-0.000	0.523
	0.218	0.219	0.001	.
Admin and support ind.	0.249	0.248	0.001	0.302
	0.432	0.432	0.001	.
Health and social ind.	0.189	0.189	-0.000	0.942
	0.392	0.392	0.001	.
Temporary help ind.	0.224	0.223	0.001	0.373
	0.417	0.416	0.001	.
Publication day	159.837	159.828	0.009	0.973
	111.102	110.887	0.285	.
	3.03e+05	3.02e+05	.	.

Notes: Columns (1) and (2) report sample averages for treated and control groups respectively. Below the means are standard deviations. Column (3) report the difference in means across treated and control group. Below the differences are the standard errors. Colmun (4) reports the p-value for zero difference. Sample: Super control vacancies from November 2021 to March 2022 are not considered.

Table C6: Balancing for vacancies: categorical variables

	Share of p-values > 0.05
Occupation (4-digit SSYK)	0.95
Industry (5-digit)	0.95
Publication week	0.97
Municipality	0.95
Commuting zone	0.94
Local labor market	0.96

Notes: Share of treated-control differences with p-value $>$0.05. Super control vacancies from November 2021 to March 2022 are not considered. There are over 400 4-digit occupations, 704 5-digit industries, 260 municipalities, 69 commuting zones. We define 138 local labor markets as combination of commuting zone and high vs. low skill groups.

C.4 Pair-level analysis

In the pair-level analysis, we collect for each experimental user the set of recommended vacancy potentially shown during their website visits. We observe a set of predetermined characteristics: X_{ij} which can vary across users i, across vacancy j or at the pair-level. We start to test balancing for pair-level variables with the following regression:

$$X_{i,j} = \gamma_i + \gamma_j + \beta T^p_{i,j} + \varepsilon_{i,j}$$

Where $D_{i,j}$ is the indicator that the pair is treated (i.e both the user and the vacancy are treated) and γ_i and γ_j are user and and vacancy fixed effect. The residuals $\varepsilon_{i,j}$ are clustered at the level of the user_id $\times$ vacancy_id. We report in Table the β estimates and the corresponding p-values. For user-level variables, we collapse the data at the user level before comparing treated and control groups. Similarly for vacancy-level variables. When variables are categorical, we run as many regressions as categories where the left-hand side variable is a dummy for that category. We report the fraction of p-values above 0.05 across all categories.

Table C7: Balancing table for variables at the pair level

Variable Names	M0	Pair treated	pval
Geographical distance to reference municipality (in km)	47.45264	-0.02720	0.47233
	158.25002	0.03785	.
Occupationnal distance to reference occupation within occupation classification	0.83810	0.00004	0.76914
	0.30462	0.00013	.
Occupational distance, measured by past transitions	0.95569	0.00000	0.94061
	0.15526	0.00006	.
Nb of days elapsed between day pub and day recom	15.30674	0.00440	0.40144
	19.96843	0.00524	.
BKM1. recom	0.13685	-0.00005	0.70693
	0.34369	0.00015	.

Table C9: Balancing table for variables at the user level

Variable Names	M0	Diff	Pval
Women	0.52548	-0.00047	0.72727
	0.49935	0.00135	.
Swedish	0.60596	-0.00236	0.07421
	0.48864	0.00132	.
Unemployed	0.52430	-0.00156	0.29342
	0.49941	0.00148	.
High-school dropouts	0.18495	0.00069	0.48705
	0.38826	0.00100	.
High-school diploma	0.38450	-0.00144	0.24824
	0.48648	0.00125	.
Post-secondary educ	0.33113	-0.00049	0.68755
	0.47062	0.00121	.
Stockholm	0.08681	-0.00021	0.76957
	0.28156	0.00072	.
Last month with positive earnings	732.59414	0.05218	0.02770
	8.01705	0.02370	.
Below median occ. distance, preexp	0.57883	-0.00191	0.13579
	0.49375	0.00128	.
Above 3rd quartiles clicks on same occ. code (SSYK), preexp	0.20585	-0.00053	0.61099
	0.40432	0.00105	.
Quartiles of search in km, preexp	2.60963	0.00272	0.38111
	1.09058	0.00311	.
Terciles of SSYK dist (occupational distance) , preexp	1.95741	0.00250	0.28461
	0.80132	0.00234	.

Table C11: Balancing table for categorical variables at the user level

Variable Names	Share
Occupation code (SSYK)	0.95
First day when device may be shown recom.	0.95
Municipality	0.95

Table C13: Balancing table for variables at the vacancy level

Variable Names	M0	Vacancy treated	pval
Quantile of popularity	3.06758	-0.00534	0.16351
	1.42000	0.00383	.
Number of vacancies	1.98953	-0.00696	0.74616
	8.31671	0.02150	.
Fixed pay	0.93668	0.00026	0.69622
	0.24354	0.00065	.
Full time	0.76500	0.00168	0.13969
	0.42400	0.00114	.
Regular employment	0.91474	0.00193	0.00987
	0.27927	0.00075	.
Experience required	0.76048	0.00043	0.70801
	0.42679	0.00115	.
Open ended	0.67874	0.00124	0.32260
	0.46696	0.00125	.

D Pair-level analysis: from empirical to theoretical quantities

In this section, we show how we estimate the decomposition of net employment effect from the user perspective within the recommended jobs.

We start by estimating the following pair-level regressions for both the application event A_{ij} and the employment event E_{ij}:

$$A_{ij} = a_{CC}(1-T_i^u)(1-T_i^v) + a_{TC}T_i^u(1-T_i^v) + a_{CT}(1-T_i^u)T_i^v + a_{TT}T_i^uT_i^v + \nu_{ij}^a \tag{10}$$
$$E_{ij} = e_{CC}(1-T_i^u)(1-T_i^v) + e_{TC}T_i^u(1-T_i^v) + e_{CT}(1-T_i^u)T_i^v + e_{TT}T_i^uT_i^v + \nu_{ij}^e \tag{11}$$

For applications, we have the following identifying relationships.

$$A_0 = a_{CC}$$
$$A_1 = a_{CC} - a_{TC}$$
$$A_2 = a_{TT} - a_{CT}$$

In these relationships, the left-hand side variables are quantities from our theoretical framework (see above) and the right-hand side variables are those in the equation (10). In the case of the number of applications, we have an additional theoretical prediction, $a_{CT} = a_{CC}$, which we can use to assess the empirical validity of our model.

Second, we consider equation (11) estimated for the outcome: employment. In this case, we have the following identifying relationships.

$$E_0 = e_{CC}$$
$$gE_0 = e_{CC} - e_{CT}$$
$$E_1 = e_{CC} - e_{TC}$$
$$(1-g)E_2 = e_{TT} - e_{CT}$$

In this case, the model is just identified: four empirical quantities to identify four theoretical quantities.

Using the empirical estimates of $A_0, A_1, A_2, E_0, E_1, E_2, g$ from Table 7, we can proceed with computing the effect of the treatment on treated and control jobs, the net employment effect, as well as the net congestion factor. These empirical quantities are displayed in Table 8.

E Extra figures

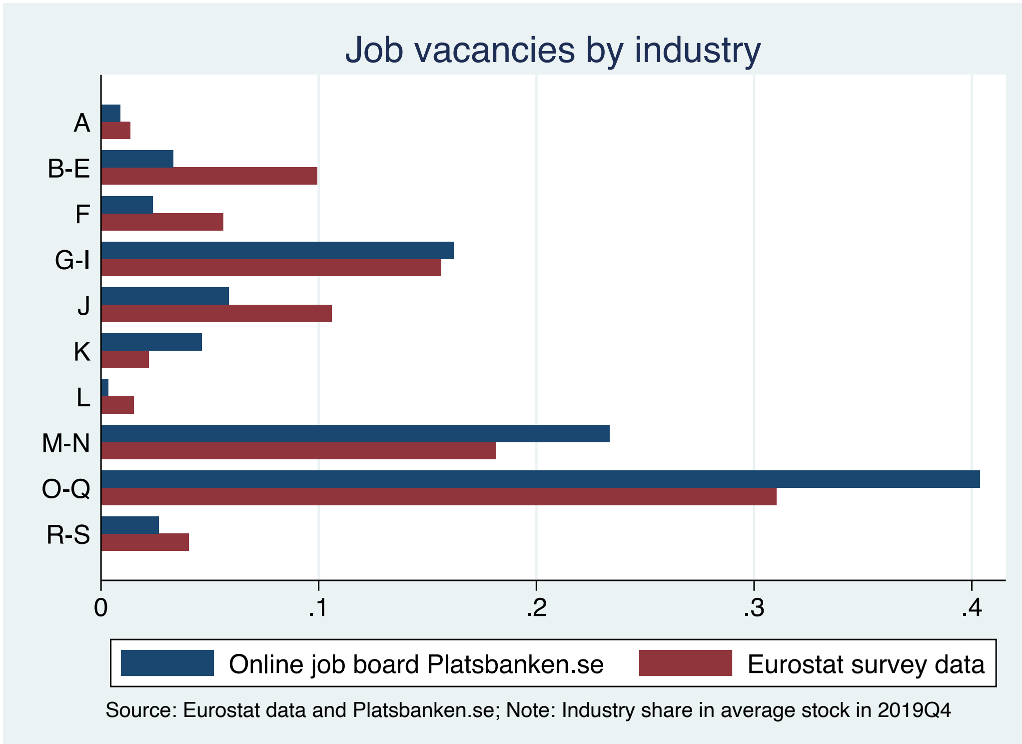


Figure F1: Vacancies on Platsbanken.se vs in Eurostat survey.

Note: We use the aggregate level of the NACE industry classification. A: Agriculture, forestry and fishing; B-E: Manufacturing (except construction); F: Construction; G-I: Wholesale and retail trade, transport, accommodation and food service activities; J: Information and communication; K: Financial and insurance activities; L: Real estate activities; M-N: Professional, scientific and technical activities; administrative and support service activities; O-Q: Public administration, defence, education, human health and social work activities; R-S: Arts, entertainment and recreation; other service activities.

In Platsbanken, we exclude vacancies from temporary help agencies, because we cannot assign them to the industry of the client firm.

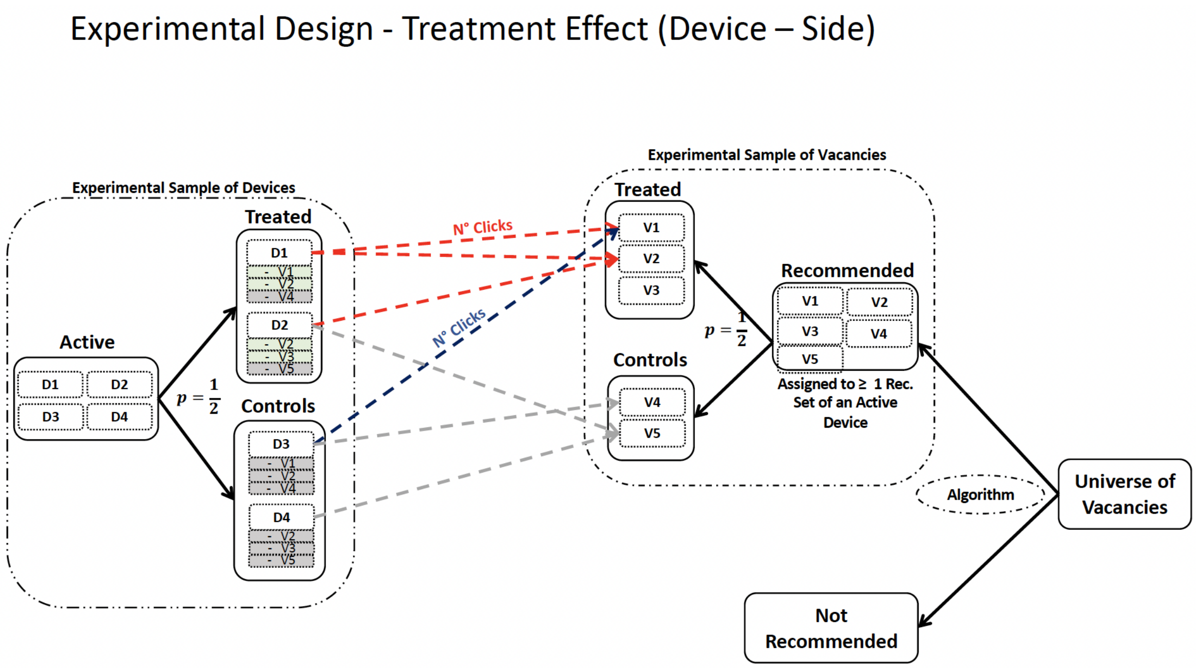


Figure F2: Experimental Design: user- and vacancy-level randomization

Note: This figure plots the time series of the monthly inflow into the experiment.

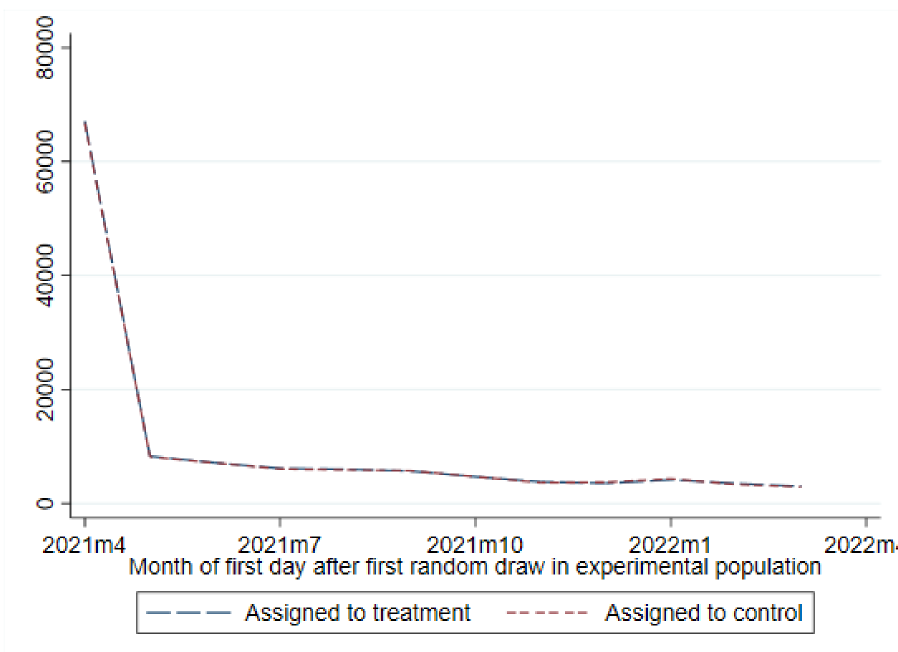


Figure F3: Monthly inflows into the experiment

Note: This figure plots the time series of the monthly inflow into the experiment. The sample includes active registered users.

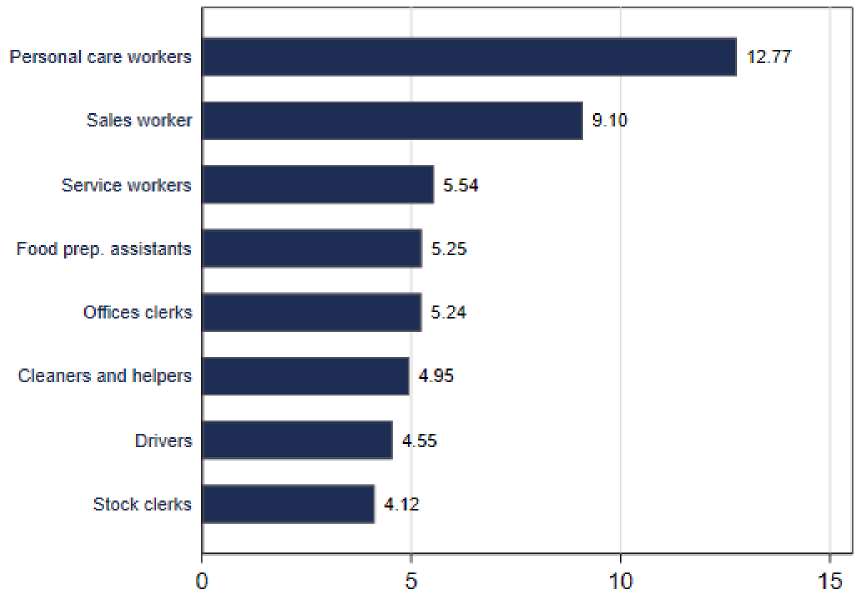


Figure F4: The eight most frequent reference occupations (2-digits)

Note: This figure shows the eight most frequent reference occupations (2-digits) in our main evaluation sample. It reports the percentage of active registered users with that reference occupation. The occupation "Service workers" includes principally "Cooks and cold-buffet managers" (26.02 %), "Waiters" (22.12 %) and "Building caretakers" (20.33 %).

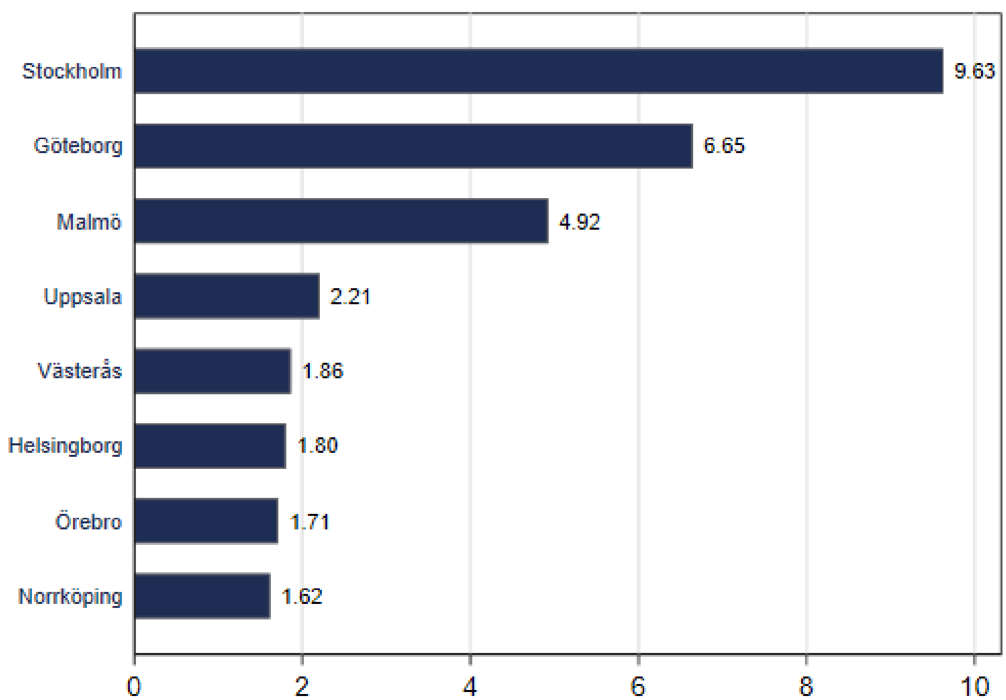


Figure F5: The eight most frequent reference municipalities

Note: This figure shows the eight most frequent reference municipality in our main evaluation sample. It reports the percentage of active registered users with that reference municipality. There are 290 municipalities in Sweden.

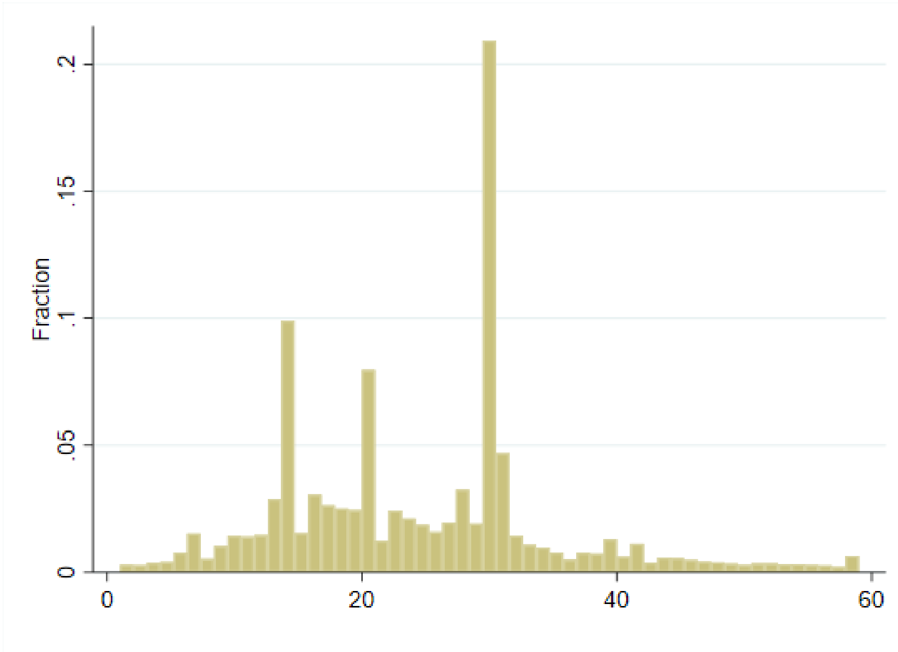


Figure F6: Number of days between publication date and application deadline, at the vacancy level

Note: This figure plots the difference between the date of publication of the vacancy, and the date of application deadline stated in the job ad. The median is 28. Sample: inflows of control vacancies ($N = 371,555$)

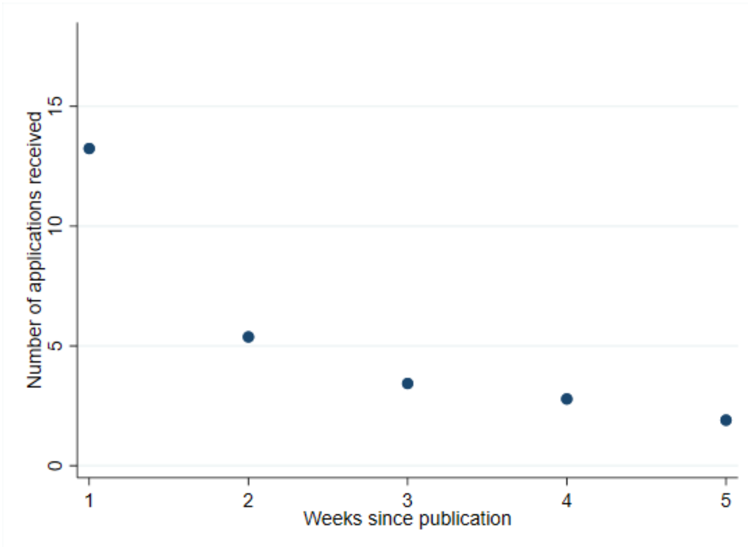


Figure F7: Mean number of applications received, by number of weeks since the day of publication

Note: This figure shows the mean of the number of applications by the number of weeks since the day of publication. It is computed on the sample of the control vacancies, for which the information related to the date of publication and the date of the deadline are not missing (N = 371,555). After the 5th week following publication, 22.63 % of the vacancies of this sample were still active.

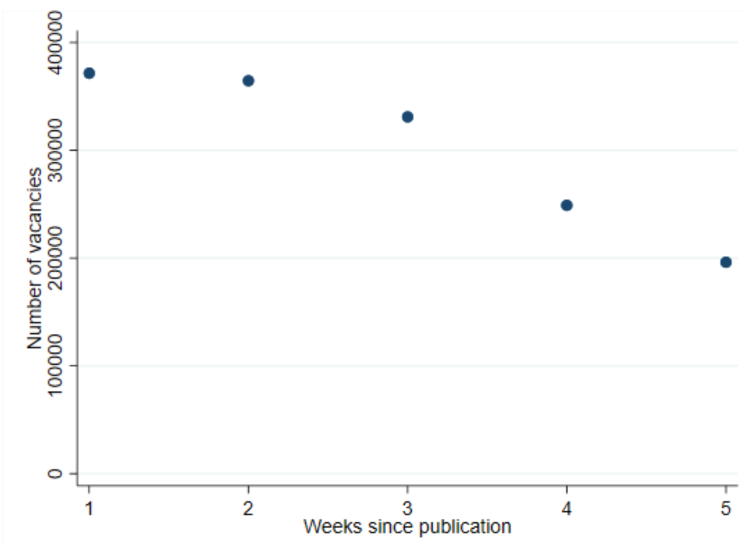


Figure F8: Number of active vacancies by number of weeks since the day of publication

Note: This figure shows the number of active vacancies by the number of weeks since publication. It corresponds to the sample used to compute each point in fiugre F7 is computed on the sample of the control vacancies, for which the information related to the date of publication and the date of the deadline was completed (N = 371,555).

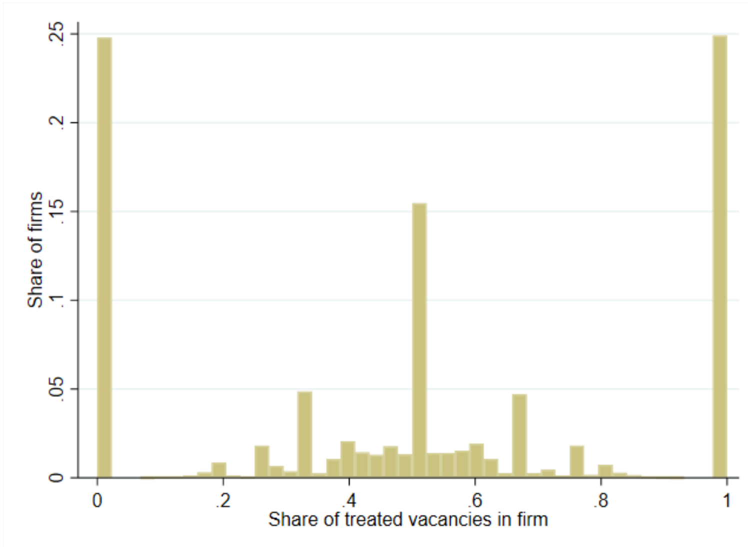


Figure F9: Distribution of the share of treated vacancies posted by firms over the experimental period

Note: This figure shows the share of treated vacancies posted by firms over the experimental period. Sample: XX firms

F Extra tables

Table T1: Applications per vacancy: Comparison with literature

Study	(1) # of applications received per vacancy (per week)	source
Our estimate	5.35	
Marinescu (2017)	7.5	CB.com
Marinescu and Rathelot (2018)	15.8/4	CB.com
Banfi and Villena-Roldan (2019)	4.06	trabajando.com
Marinescu and Wolthoff (2020)	59/(16/7)	CB.com

Note: this table reports estimates of average number of applications received per vacancy per week. Our estimate is computed from the inflow of control vacancies during the experimental period. In each week between the publication date and application deadline, we count the number of application received (can be 0). We censor weeks that are more than six weeks after publication. Note that more than 50% of vacancies are no longer available for applications by week 6 (see Figure F8). We then report in the table the simple average.
Marinescu (2017) "On average, each vacancy receives about 30 applications per month" (comment of table 1, p73).
Banfi and Villena-Roldan (2019): we take the ratio of average number of ads (Table 2) by the period for which firms pay for posting (60 days).

Table T2: Treatment effect on probability to click on a given day

	(1) Active	(2) Active
User is treated	0.00027	0.00052
	(0.00022)	(0.00063)
Sample	All	Registered User
Observations	178,091,554	26,688,448
Control mean	0.082	0.11
Outcome sd	0.27	0.31
Pct impact	0.33	0.49

Note: This Table reports the treatment effect on the active status of users, i.e. whether they click on at least one vacancy in a given day. The regression is at the user X day level. We consider all users who are included at least once in the training set of the recommender system in Column (1), and restrict the sample to registered users in Column (2). The sample comproses all the days when recommendations are generated for those users. We create a dummy variable indicating whether the user views at least one vacancy during that day, and we denote it active status. For job seeker i on day d, we run the following regression: $Y_{id} = \alpha + \delta T_i^u + \epsilon_{id}$ where standard errors are clustered at the user level.

Table T3: Treatment effect on the number of days when vacancy is shown as a recommendation to at least one device

	Days in recommendation set
Vacancy is treated	0.0342
	(0.0600)
Observations	6.05e+05
Control mean	18.85
Outcome sd	18.12

Note: This Table reports the treatment effect on the the number of days when vacancy is shown as recommendation to at least one device. The regression is at the vacancy level. We consider all vacancies that are included at least once in the recommendation set of an experimental user (either control or treated). The dependent variable reports the number of days in which the vacancy is included in at least one recommendation set. For vacancy i, we run the following regression: $Y_i = \alpha + \delta T_i^v + \epsilon_i$, with robust standard errors.

Table T4: Effect on daily clicks and daily applications per **registered** user

	(1)	(2)	(3)	(4)
	All	Recommended jobs		
	jobs	Yes	Yes but not shown	No
		Panel A: Clicks		
User is treated	-0.0255*	0.0632***	-0.00172**	-0.0870***
	(0.0139)	(0.00118)	(0.000833)	(0.0136)
Control mean	4.021	0.115	0.116	3.790
Outcome sd	3.698	0.386	0.387	3.563
pct impact	-0.635	55.05	-1.484	-2.296
		Panel B: Applications		
User is treated	-0.00996	0.0118***	-0.000628	-0.0212***
	(0.00691)	(0.000489)	(0.000399)	(0.00646)
Control mean	0.880	0.0315	0.0316	0.817
Outcome sd	1.562	0.195	0.195	1.482
pct impact	-1.132	37.56	-1.986	-2.593
		Panel C: Within-day conversion rates (in %)		
Control users	21.9	27.5	27.3	21.5
	(0.13)	(0.28)	(0.28)	(0.13)
Treated users	21.8	24.4	27.2	21.5
	(0.13)	(0.27)	(0.29)	(0.14)
Observations	2,839,834	2,839,834	2,839,834	2,839,834
no of users	257147	257147	257147	257147
no of individuals	182529	182529	182529	182529

Sample: active registered users (i.e., who click at least once on day d)

Note: this table reports the treatment effects from user-day level regressions. For each user, we consider all post-randomization days when she clicks on at least one vacancy. In Panel A, we report treatment effects on the daily number of clicks per user (column 1), restricting to clicks on treated vacancies in users recommendation set in Column (2). Column (3) restricts to clicks on control vacancies in users recommendation set, which are not shown. Column (4) considers clicks on the complement sample of vacancies (user-specific complement). We report treatment effects both in absolute value and in percentage (as the ratio of absolute effect on the corresponding average among control users). In Panel B, we report treatment effects on daily applications. In Panel C, we compute conversion rates from clicks to applications as implied by panels A and B.

Robust standard errors clustered at user level in parenthesis. They are computed with delta method in Panel C and E. The conversion rate is obtained as the ratio of the application estimate ($\hat{A}$) and of the click estimate ($\hat{C}$). Let us denote $se(C)$ (resp. $se(A)$) the standard errors of $\hat{C}$ (resp. $\hat{A}$). Then using the delta method, we obtain the standard errors of the ratio ($\hat{R} = \hat{A}/\hat{R}$) as: $se(R) = \left(se(A)^2 / \left(\hat{C}\right)^2 + se(C)^2 \left(\hat{A}\right)^2 / \left(\hat{C}\right)^4\right)^{1/2}$.