Caseworker versus Algorithmic Prediction of Jobseeker Success *

Lovisa Rambjer †

March 19, 2025

Preliminary and incomplete

Abstract

Statistical profiling of jobseekers, designed to distinguish jobseekers at risk of long-term unemployment from those likely to secure employment quickly, are being adopted in public employment services across OECD countries. These tools guide the allocation of active labor market programs worth billions of dollars annually. Despite their growing prevalence and potential impact, little is known about the real-world performance of statistical profiling tools — and even less about how they compare to human decision making. While algorithms often outperform humans in prediction tasks, the conditions under which human discretion improves or undermines their effectiveness remain largely unexplored. This paper examines the use of an algorithmic prediction tool at the Swedish Public Employment Service. Leveraging random assignment of jobseekers to caseworkers and variation in caseworker propensity to assign jobseekers to independent job search, I reconstruct the counterfactual scenario of strict algorithmic allocation and estimate employment outcomes. This approach allows for an unbiased comparison of caseworker versus algorithmic prediction of jobseeker reemployment probability. Preliminary results reveal that caseworker deviations from the algorithm are split evenly between improving and worsening predictions, resulting in a zero net effect on conditional employment rates. These findings highlight the need to understand when human judgment complements or counteracts algorithmic decision-making.

*I would like to thank Georg Graetz and Peter Fredriksson for their invaluable guidance and support. I am also grateful for helpful suggestions from Will Dobbie, Simon Ek, Daniele Paserman, Ashesh Rambachan and Johan Vikström, as well as comments from seminar and conference participants at the Boston University Micro workshop 2025, UCLS Annual Meeting 2024 (Uppsala), the Swedish Public Employment Services, and at the Nordic Public Policy Symposium 2024 (Copenhagen).

†Uppsala University and UCLS. Email: Lovisa.Rambjer@nek.uu.se

1 Introduction

Active labor market policy (ALMP) plays a crucial role in labor market policy and the unemployment insurance system in many developed economies. They aim to mitigate the disincentive effects of unemployment benefits and address labor market frictions. ALMP also represents a substantial public expenditure — on average, OECD countries allocate 0.5 % of GDP to these programs.[1] As unemployment rises and resources remain limited, the effectiveness of ALMP depends not only on the interventions themselves but also on *who* receives them. To support caseworkers in making assignment decisions and improve the targeting of ALMP, statistical profiling tools have become widely adopted across OECD countries. These tools aim to distinguish jobseekers at risk of long-term unemployment from those likely to secure employment quickly on their own. The idea of statistical profiling for jobseeker targeting dates back to the 1990s, with the U.S. introducing statistical profiling for customer segmentation as early as 1996.(OECD, 2019) Today, advancements in machine learning have enabled the development of more sophisticated and high-performing algorithms. However, despite their widespread adoption, we still know little about how these algorithms perform in practice — or how they compare to human decision making.

Despite increasing reliance on algorithmic profiling and significant policy implications, rigorous evidence on its real-world effectiveness remains scarce. The existing literature is sparse (e.g. Ernst et al. 2024 and 2023 van den Berg et al.), and often suffers from significant methodological challenges. Three key limitations stand out. First, observed outcomes are endogenous to prior caseworker decisions. There are systematically missing data on outcomes in the absence of treatment, known as the selective labels problem (Kleinberg et al., 2018). Algorithms are typically evaluated on outcomes of either the non-random sample of untreated individuals, or on outcomes affected by treatment. This limits the informativeness of conventional machine learning accuracy measures for assessing their effectiveness in targeting ALMP. If not addressed, the problem of endogenous outcomes will bias any comparison between the predictive performance of caseworkers versus an algorithm. Second, standard accuracy metrics may not align with policymaker priorities. In the public employment service (PES) context, identifying jobseekers at the highest risk of long-term unemployment — those at the bottom of the distribution — may be far more important than optimizing average prediction accuracy. Finally, we lack a clear understanding of how caseworkers interact with the algorithmic recommendations when making final decisions. Do they improve upon algorithmic predictions using private information, or do they introduce bias and noise into the decision-making process?

[1] OECD 2021, https://goingdigital.oecd.org/en/indicator/42

This paper investigates the impact of human discretion over algorithmic recommendations in the context of assignment to job search assistance at the Swedish Public Employment Service (PES). Using data from PES matched with Swedish registry data, I study an algorithm introduced in 2020 that predicts jobseekers' reemployment probabilities to assess eligibility for job search assistance. In particular, I examine the implications of caseworker deviations from the assignment recommendations. The key question is whether such deviations reflect valuable private information about jobseekers — such as effort or motivation — not captured by the algorithm but potentially observable to caseworkers, or whether caseworker deviations add back human bias and noise into the process. Will caseworker discretion improve the prediction of job search success at the Swedish public employment services? By leveraging random assignment of jobseekers to caseworkers and variation in caseworker propensity to assign jobseekers to independent search, I deal with the selective labels problem, and can estimate the latent employment outcomes under strict adherence to the algorithm. This allows me to compare the predictive performance of the algorithm to the implicit predictive performance of the caseworkers.

At the Swedish PES, job search assistance consists of a six- to twelve-month program, supplied by private providers in a competitive market. The algorithm was introduced to ensure that job search assistance and is targeted at jobseekers with the lowest probability of finding employment independently. Evaluating the performance of the algorithm and caseworkers requires understanding the extent to which individuals assigned to job search assistance would have been able to find a job on their own. However, outcomes for these individuals in the absence of assistance remain unobservable. We will only be able to observe the outcomes of jobseekers who the caseworkers have assigned to independent job search, such that there is non-random missing data on outcomes.

This selective labels problem constitutes the main empirical challenge in making an unbiased comparison of the predictive performance of caseworkers versus the algorithmic counterfactual using observational data. A traditional way of dealing with this selection problem would be to control for some jobseeker characteristics and make an assumption that conditional on observables, caseworker decisions are uncorrelated with the employment potential, i.e. they are as good as random. An assumption which is particularly problematic in this project, where the purpose is to investigate whether caseworkers use private information to improve on the prediction. To address this, I employ a quasi-experimental method which exploits random assignment of jobseekers to caseworkers as well as variation in caseworkers' assignment rates—that is, their propensity to assign jobseekers to independent job search.

This approach enables the evaluation of predictive performance by focusing exclusively on the employment outcomes of jobseekers assigned to independent job search. Crucially,

this method does not rely on assumptions about the efficacy of job search assistance. By estimating the employment rate of individuals assigned to independent job search while holding assignment rate constant, I can directly compare the predictive performance of caseworkers and the algorithm. Evaluating performance at a fixed assignment rate isolates differences in accuracy while holding both budgetary constraints and caseworker preferences constant. The method builds on bounding (Kleinberg et al., 2018; Rambachan et al., 2024) and extrapolation approaches (Angelova et al., 2023; Arnold et al., 2022), previously used to study human and algorithmic decision making. At the PES, most jobseekers get assigned to independent search, meaning that there will be a high share of caseworkers with very high or 100 % assignment rates. The high assignment rates facilitates a precise, and unbiased, estimate of the algorithmic counterfactua[2]

Preliminary results reveal that caseworker deviations from the algorithm are split evenly between improving and worsening predictions. On average, caseworkers' deviations yield the same conditional employment rate as strict algorithmic adherence. But there is substantial performance heterogeneity across caseworkers. 52 % of caseworkers perform worse than the pure algorithmic allocation of jobseekers, and 49 % of caseworkers perform worse than a random allocation of jobseekers. And both high and low performing caseworkers are deviating to the same extent, implying that discretion is not necessarily based on superior private information.

At face value, the results suggest that many caseworkers are making systematic errors in predicting jobseekers' reemployment probabilities. To facilitate a correct interpretation of the results, I must formalize the conditions under which these results can indeed be interpreted as proof of misprediction. An alternative explanation might be that caseworkers use a different rule to allocate jobseekers. One such rule could be that they assign jobseekers based on predicted return to job search assistance, which could lead to a different allocation decision if predicted return is not decreasing in reemployment probability for all jobseekers. Such a decision rule would, however, be in conflict with the aim of the policymaker who introduced the algorithm to ensure that labor market interventions are allocated based on jobseekers' need for support. This relates to a broader discussion on diverging objective functions between algorithms and human decision makers, e.g., Ludwig and Mullainathan (2021) and Rambachan (2024). Understanding these differences is critical to inform human oversight policy as the use of predictive algorithms to aid in decision making becomes more widespread.

The question of human discretion over algorithmic recommendations extends beyond the

[2]Note: In the current version of the paper, and the preliminary results presented, the extrapolation method is used.

PES setting. Machine learning algorithms are increasingly used to assist human decision making across a range of domains such as healthcare, criminal justice, and credit scoring. By processing vast amounts of data, algorithms excel in solving complex prediction problems where humans may be limited by prejudice or simply inability. And as opposed to human decision makers, algorithms operate under clearly defined and controllable objective functions. However, while much speaks in favor of algorithms outperforming humans, their actual effectiveness depends on how humans interpret and act on their outputs. A fundamental challenge in algorithmic implementation is what determines, and how we can facilitate, successful human-algorithm collaboration (Ludwig and Mullainathan, 2021). This depends on human ability both to make use of private information and interpret the algorithm's signals.

This study contributes to this discussion by examining a unique setting where caseworkers must explicitly justify deviations from algorithmic recommendations. And it turns out that the vast majority of deviations occur when jobseekers claim they are already working part-time or have a pending job offer. And by leveraging a shift in the algorithmic regime, I can shed light on the importance of predictive performance as well as the design of the algorithmic output.

In summary, this project makes two key contributions. First, it is the first to provide evidence on the impact of caseworker discretion over algorithmic recommendations in a PES context. Including perspectives such as fairness in assignment and examining caseworker preferences. Second, leveraging rich administrative data and institutional features, it provides novel insights into mechanisms determining the success of human-algorithm collaboration. By shedding light on these dynamics, this study offers insights for improving algorithmic governance in labor market policy and related fields.

1.1 Literature review

This project is related to two main strands of literature. Firstly, this paper contributes to the growing literature at the intersection of AI and economics, evaluating the societal impact of AI tools and the scope of using machine learning to improve prediction in prediction policy problems. Researchers have long recognized the central role of prediction in human decision-making across domains such as medicine and criminology (Ohlin and Duncan, 1949; Dawes et al., 1989). A well-documented finding is that human predictions are often noisy and bias, creating the potential for substantial gains from statistical or rule-based decision aids. (Dawes, 1979; Grove et al., 2000).

The modern literature on algorithmic decision-making has focused extensively on high-stakes contexts such as pretrial hearings (Kleinberg et al., 2018; Angelova et al., 2023),

and medical diagnosis (Mullainathan and Obermeyer, 2022; Chan et al., 2022). In the labor market setting, research has examined the effectiveness of algorithmic hiring tools. For example, Cowgill (2020) studies resume-screening technologies through an experiment where algorithmic recommendations randomly overrule human decisions. He finds that algorithmic resumee screening produces candidates who are more likely to receive and accept job offers compared to those selected by human decision-makers. Similarly, Hoffman et al. (2018) exploits a gradual roll-out of job testing technology across establishments and find that it improves hiring outcomes. Their study also reveals a negative correlation between managerial deviations from algorithmic recommendations and the quality of hires. However, due to the selective labels problem, the study cannot provide causal estimates of the impact of managerial discretion.

In the PES context, research has examined the effectiveness of different allocation mechanisms for labor market programs. Lechner and Smith (2007) analyze jobseeker assignments in Switzerland and find that caseworker allocation performs no better than random assignment, while statistical profiling based on predicted treatment effects could significantly improve employment outcomes. Their analysis relies on non-experimental methods and assumes selection on observables. More recent studies have examined algorithmic prediction tools. Ernst et al. (2024) investigate how algorithmic risk profiling affects the selection into job search assistance in the Flemish PES, Belgium. While van den Berg et al. (2023) combine survey and administrative data from Germany to compare the predictive performance of reemployment probabilities across jobseekers, caseworkers, and an algorithm.

This paper builds on and extends the existing literature in three key ways. In contrast to previous literature in the labor market and PES settings, this paper employs a quasi-experimental approach that explicitly addresses the selective labels problem, enabling the estimation of counterfactual jobseeker outcomes under algorithmic assignment recommendation. This study is one of the first to provide evidence on how caseworkers interact with algorithmic recommendations in a real-world labor market setting. By examining when and why caseworkers override algorithmic recommendations, this paper contributes to the broader discussion on algorithmic governance and the conditions under which human discretion improves or undermines algorithmic predictions.

Secondly, this paper contributes to the literature on the role of caseworkers at the PES. Existing research has examined the matching of jobseekers with caseworkers based on characteristics (e.g., Behncke et al. (2010)) and work strategies (e.g., Arni et al. (2022)).Cederlöf et al. (2021) leverage as-if random assignment of jobseekers to caseworkers to estimate caseworker value-added and find that caseworkers matter for time in unemployment as well as job quality. Their findings highlight experience and work strategies as critical for caseworker

performance. This paper extends the literature by investigating heterogeneity in caseworkers' prediction performance, e.g. by caseworker experience.

The remainder of the paper is structured as follows: Section 2 provides an overview of the institutional context and introduces the data. Sections 3 and 4 describe the conceptual framework and empirical strategy, respectively. Finally, the results are presented in Section 5.

2 Data and Institutions

2.1 The Swedish Public Employment Services

2.1.1 Introduction of the algorithmic profiling instrument

The Swedish public employment services has undergone extensive organizational reformation in recent years. In 2010, the Act on Systems of Choice (LOV) within the Public Employment Services was passed which enabled privatization of active labor market policy. [3] In 2014 the first program with private providers of job search assistance was introduced. Finally, a political agreement in January 2019 established that the PES mandate would change fundamentally to focus on monitoring of private providers, statistics and analysis as well as provide digital infrastructure, while job search assistance would be provided mainly by private providers on a competitive market. [4] The main task of PES caseworkers would be to make labor market assessments and determine eligibility to job search assistance and active labor market programs. The reformation coincided with substantial budget cuts, resulting in fewer offices and centralization of service provision.

As part of the reformation of the PES services, a new program of privately provided job search assistance, matching services, was introduced in 2020. The program, Krom, was piloted in 2020 and the roll-out was completed in the end of 2021. The PES also introduced an algorithm, i.e. a profiling instrument, predicting reemployment probability to aid caseworkers in the labor market assessment. To enable evaluation of the new program, the algorithmic recommendation was randomized for jobseekers close to the thresholds, see Egebark et al. (2024) for the evaluation. During recent years the matching services has been the largest labor market program at the Swedish PES, with more than 70 000 participants during 2022. (Egebark et al., 2024)

In April 2023 Krom was replaced by a similar program, ROM2, still consisting of privately provided matching services. For the purpose of this paper, there were three main

[3] Proposition 2009/10:146 Valfrihetssystem hos Arbetsförmedlingen

[4] The January agreement https://www.socialdemokraterna.se/download/18.1f5c787116e356cdd25a4c/1573213453963/Janua

differences between the programs. Firstly, while the Krom-algorithm did not include time of registration in the prediction model, the ROM2-algorithm is a survival model which take time of registration into account, resulting in a better prediction. Secondly, the output from the Krom algorithm is simply an assignment recommendation, while the recommendation from the ROM2-algorithm is complemented with an ranking of the most important variables in predicting the reemployment probability. Thirdly, within Krom the caseworkers were instructed to only deviate in one direction — they were strongly recommended to not assign a jobseeker to the macthing services unless the algorithm recommended so. This resulted in fewer participants in the matching services than desired, which led to an update of the system. Within ROM2 caseworkers can freely deviate from the algorithmic recommendation in both directions.

Three regimes of matching services

1. March 2020 - December 2021: Pilot and role out of Krom.
2. January 2022 - April 2023: Krom program was running nationwide.
3. April 2023 - today: New program ROM2 replaced Krom.

Which regimes to study is a trade-off between sample size, precision and follow-up period on one hand, and external validity on the other. I will exclude the pilot and role out period for two main reasons. Firstly, the period was characterized by the covid-19 pandemic, which had an extensive impact on the labor market. Secondly, it is likely that caseworkers change their response to the algorithmic recommendation over the initial learning period. I want to estimate the impact of caseworker versus algorithmic decision-making during stable conditions. Thus the analysis will focus on the second and third regime.

2.1.2 Labor market assessment and assignment decision

Caseworkers make the final assignment decision based on (1) the algorithmic recommendation, i.e. the output from the profiling instrument, and (2) private information gathered in meeting with the jobseeeker. The caseworkers are generally advised to follow the algorithmic recommendation. The purpose of the algorithm is to ensure that jobseekers are assigned to different treatments based on their need of support, measured as their reemployment probability. The algorithmic recommendation, and final assignment, can be divided into three main categories ordered from highest to lowest reemployment probability. (Arbetsförmedlingen, 2021)

1. **Independent job search**
   Searching for a job mostly independently, with the help of digital self-service tools.

2. **Job search assistance**
   Privately provided job search assistance including coaching and guidance, CV writing, interview training, and job referrals.

3. **Intensive support**
   For jobseekers in need of more extensive and intense support such as labor market training, labor market education, language training.

2.1.3 The algorithm

The aim of the profiling instrument is to make a prediction on the probability that a jobseeker becomes employed on their own, or starts studying, within six months. The algorithm takes individual characteristics such as education, search-occupation, country of birth, age and disability into account, as well as zip code level variables such as unemployment rate and median income. See Table A1 for full list. The information comes from both PES registries and the jobseeker's self-reported data. From April 2023, the prediction model was updated to better handle time of registration, enabling the inclusion of time of registration into the model as well as a larger and more representative training population. While the old model was trained on only newly registered jobseekers, the new model is trained on the stock of jobseekers.

The performance of the new algorithm is substantially better than the old. The old algorithm had an *accuracy*, i.e. a share of correctly classified observations, of 68 %. The new algorithm has an accuracy of 74.9 %. A random classification would generate an accuracy of 57 %. Another measure of performance is the ROC-AUC, which measures how well an algorithm rank the jobseekers in terms of their reemployment probability.[5] This is a standardized measure which can be more easily compared across models. The AUC of the new algorithm is around 0.8, which is considered a well performing algorithm. 0.5 is equivalent to a random assignment. The calculations are done using realized outcomes without taking into account that some individuals has been subject to interventions and assistance at the PES. Under the assumption that the PES assistance and interventions improve the labor market outcomes of jobseekers, the PES argues that these measures should be viewed as conservative. (Arbetsförmedlingen, 2023)

[5] Receiver Operating Characteristic - Area Under the Curve gives the probability that the model will rank a randomly chosen positive evaluated observation higher than a randomly chosen negatively evaluated observation. It ranges between 0 and 1, with 0.5 indicating a performance in line with random allocation.

2.1.4 Process for newly registered jobseekers

The PES caseworker organization is divided into two business areas, the *Direct/Instant* and *Local* services. The two cater to somewhat different populations of jobseekers, where business area Direct offer *Personal Distance Meeting* (PDM) services, that is remote meetings, to the whole country. I will refer to these offices as central offices. The Local offices provide services to a restricted geographical region. Meetings could be either remote or in person.

The process for newly registered jobseekers is standardized and applies to all jobseekers with some exceptions. The standard procedure is to register with the PES online and book an appointment for a first assessment with a caseworker at a central office. The meeting will be digitally, over call or video call. If the jobseeker cannot register online, for example because they do not have an electronic ID or because they need a translator to participate in the meeting, they will register at a local office. At the online registration, the jobseeker gets profiled the first time. In meeting with the caseworker, the profiling is updated if any new information is revealed. The caseworker then makes an assignment decision based on the algorithmic recommendation and private information received in the meeting. The caseworker can also decide to refer the jobseeker to a local office if an in-person meeting is needed in order to make an assessment. Thus on average, jobseekers assisted at the local offices are expected to have lower reemployment probability.

In the standard case there is no matching between jobseekers and caseworkers at any stage of the process and the jobseekers do not get a personal caseworker. Each time the jobseeker contacts the PES, they will be assisted by a random caseworker.

Young jobseekers between 16-24 years old, jobseekers who have previously been long term unemployed and have participated in specialized programs and jobseekers with an unidentified occupation are exempted from the standard procedure and will meet with specialized caseworkers.

To ensure random assignment of jobseekers to caseworkers, I will only include jobseekers who are subject to the standard procedure at the central offices.

Summary of the process[6]

1. **Registration**

    - The jobseeker signs up at the PES to claim unemployment insurance. The algorithm takes self-reported data as well as data from the PES-registries as input and provides an algorithmic recommendation.

[6] Based on interviews with PES staff and internal document - Memo Gemensam process för beslutshandläggare – KROM och STOM, Arbetsförmedlingen 2020.

- The jobseeker signs up for a meeting with a caseworker. There is no matching between jobseekers and caseworkers.

2. **Meeting and assignment decision**

   - The caseworker makes an assignment decision. The decision is based on a labor market assessment which relies on the algorithmic recommendation and the jobseeker's answers to some complementary questions and other private information. If an in-person evaluation is needed in order to make an assignment decision, the caseworker can refer the jobseeker to a local office.
   - The caseworker sets up an activity plan, with activities based on the assignment decision. If assigned to job search assistance, the jobseeker picks a private provider for the services.

3. **Follow-up and conclusion**

   - The assignment is continuously assessed by processing documents and signals from both the private provider and the jobseeker. During follow-up, the caseworker can make a decision to initiate or prolong the assignment to JSA. The case is closed when the jobseeker has found a job or started an education.

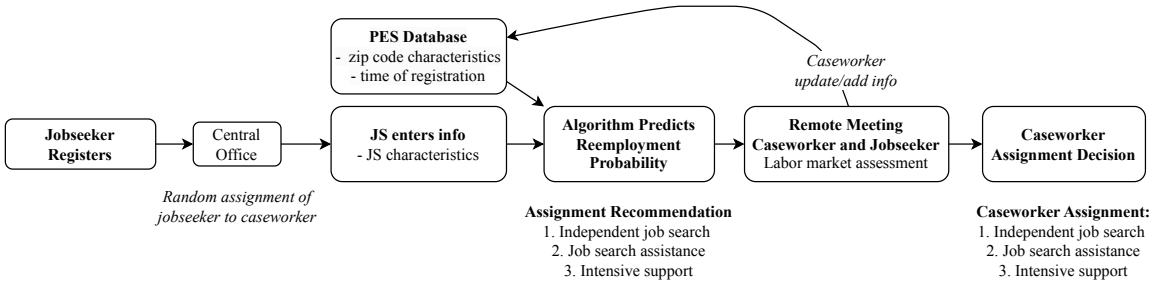
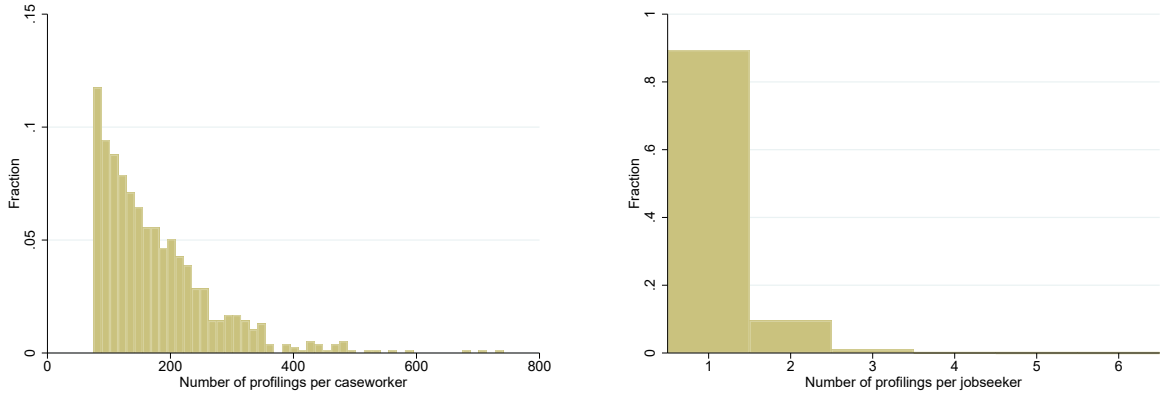


Figure 1: Process newly registered jobseekers

Note: Process for newly registered jobseekers. Randomization of algorithmic assignment recommendation for jobseekers close to the cutoffs between the three classes. Caseworkers need to enter a reason for deviating when deviating from job search assistance to independent job search or intensive support.

2.2 Data

I have data on the universe on profilings from the introduction in 2020 until today. On the jobseeker side I have access to detailed PES data linked to Swedish registry data. Regarding



Number of profilings per caseworker

Number of profilings per jobseeker

Figure 2: Number of profilings per caseworker and jobseeker

Note: Final sample, including only caseworkers who are handling at least 75 profilings per year.

caseworkers I have limited data on individual characteristics from the personnel registries at the PES. For this version of the paper, I am using data on jobseekers registered and profiled from 1st of January 2022 to the last of December 2022.[7]

The analysis focus on the binary classification problem of assignment to job search assistance and independent job search. Thus, all jobseekers for who the algorithm recommends intensive support and/or who the caseworker assigns to intensive support will be excluded from the analysis.

2.2.1 Sample selection

The data includes jobseekers registered from January 1 2022 and one spell per jobseeker. The sample selection is done based on the institutional features at the PES, and will include jobseekers who are subject to random assignment to a caseworker. This excludes jobseekers who are below 25 years old, who have previously had and have the right to re-enter the job- and development guarantee, and who are missing a target occupation. The analysis only include decisions made at the central offices, as the randomization is more likely to hold for those cases. In the final analysis sample there are 136,392 proflings, 121,968 jobseekers and 772 caseworkers.

In the final sample, caseworkers are handling at least 75 profilings. And the vast majority of jobseekers are profiled one or two times. See Figure 2 for the distribution of profilings across caseworkers and jobseekers.

[7]Analysis using data from the period after April 2023 will be added later.

2.2.2 Data on profiling score

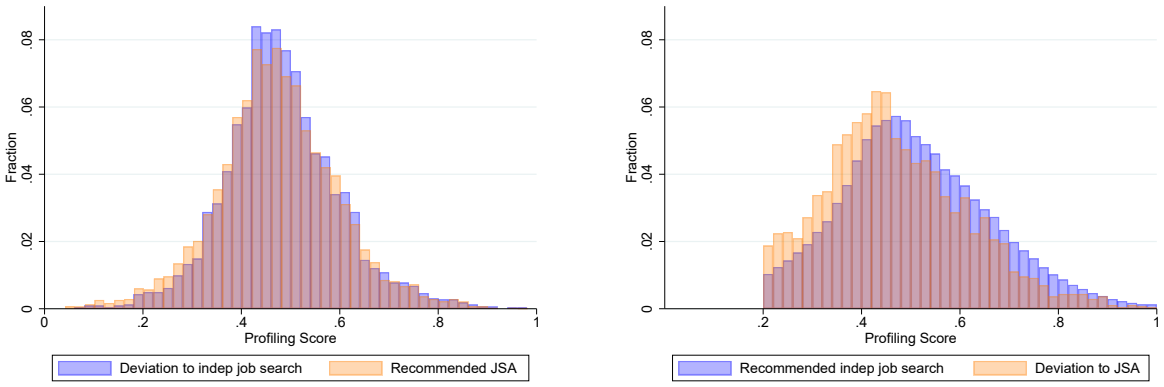
During 2022, the algorithmic recommendation was a function of profiling score and time of registration at the PES. Where the cutoffs for the algorithmic recommendation varied with enrollment duration. Importantly, the caseworker cannot observe the profiling score, but only observes the final algorithmic recommendation. Figure 3 shows the distribution of scores across the four different algorithmic classes. A feature of the system during 2022 was that in order for a newly registered jobseeker to be recommended JSA, they needed very low profiling scores. As a result, the distribution of scores of jobseekers recommended independent search is very wide.



Figure 3: Profiling score and algorithmic recommendation

Note: The fraction within each category sums to 1. Number of observations: independent search 128,836; JSA track 1 : 3,097; JSA track 2: 3,670; JSA track 3: 780

Figure 4 shows the distribution of scores for both algorithmic recommendations and caseworker deviations. Among jobseekers recommended for job search assistance, caseworkers deviated to independent search across the entire distribution of scores. For those recommended independent search, the distribution of scores for deviations is slightly shifted to the left. Overall, caseworker deviations occur across the full range of scores. This could either reflect that caseworkers are making a different labor market assessment compared to the algorithm, or that caseworkers are making noisy decisions. It remains unclear whether a differential labor market assessment would stem from (i) a different evaluation function based on public information, (ii) access to private information.



(a) Job search assistance

(b) Independent job search

Figure 4: Profiling score by algorithmic recommendation

Note: The fraction within each category sums to 1. Number of observations: algorithmic recommendation independent job search: 128,836; caseworker deviation to JSA: 2,722; algorithmic recommendation JSA: 7,556 ; caseworker deviation to independent job search: 3,228.

2.2.3 Data on deviation

Table 1 summarize the algorithmic recommendations and caseworker assignment decisions. The main deviation event is at the time of profiling, where caseworkers can make a decision to deviate from the algorithmic recommendation of assignment to job search assistance. Caseworkers are strongly recommended to not deviate in the other direction, i.e. from independent job search to job search assistance, explaining why there are proportionally few deviations in that direction. Out of the 7,556 cases recommended job search assistance, caseworkers deviated in 43 %. And out of 126,114 cases recommended independent job search, caseworkers deviated in 2 %. Resulting in a total deviation rate of 4 %.

Table 1: Algorithmic recommendation and caseworker assignment

		Caseworker assignment	
		Job search assistance	Independent job search
Algorithmic recommendation	Job search assistance	4,328	3,228
	Independent job search	2,722	126,114

Figure 5 shows the distribution of deviation rates across caseworkers.

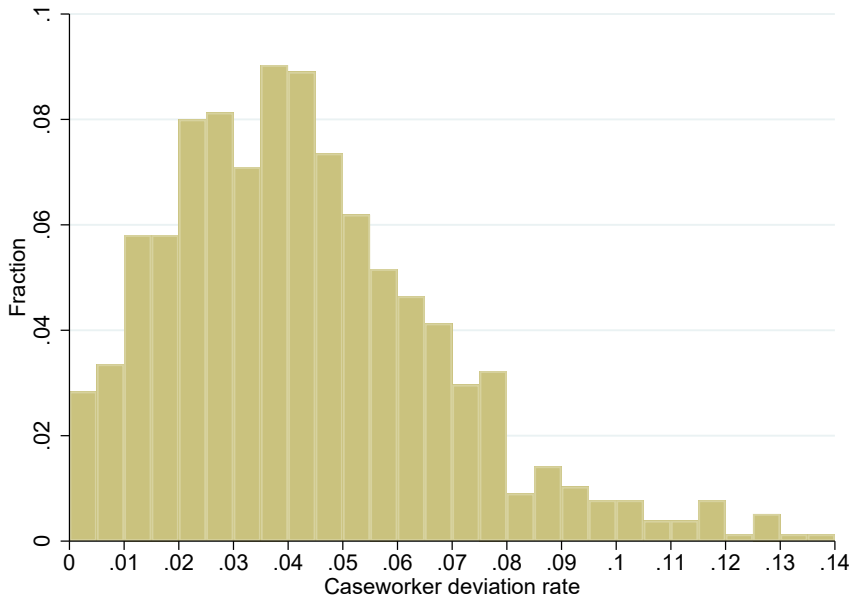


Figure 5: Caseworker deviation rate

Note: Deviation rate at caseworker level.

3 Conceptual Framework

The algorithmic recommendation and caseworker assignment decision can be illustrated using a simple theoretical framework. The framework, as well as the empirical analysis, focuses on the margin of assignment to independent job search and job search assistance, simplifying the problem to a binary classification. The setup builds on the general framework developed by Angelova et al. (2023).

Policymakers have determined that assignment to labor market interventions at the Swedish PES should be based on jobseekers' need for support as measured by reemployment probabilities. This is to ensure cost-effective use public resources and to avoid crowding out, such that assistance is not provided to a jobseeker who would have found employment independently. These considerations are particularly important in a system where private providers' compensation depends on jobseeker outcomes rather than the services delivered. (Arbetsförmedlingen, 2021).

The algorithmic recommendation is an important tool to fulfill that goal. I will assume that caseworkers have internalized the aim of the organization, and are assigning job-seekers to treatment based on their reemployment probability. An assumption which could later be relaxed to allow for caseworker payoff functions to differ from the policymaker's. Strictly evaluating the accuracy of the prediction, I will focus on the sample of jobseekers who are

assigned to independent job search. For this subsample, jobseeker outcomes and prediction accuracy will coincide - the employment rate will be the same as the accuracy rate. Comparing performance at a given assignment rate will allow me to estimate the difference in accuracy, holding preferences constant.

3.1 Setting up the assignment problem

Let there be a set of caseworkers j making a binary assignment decision $D_{i,j} \in \{0, 1\}$ on a population of job-seekers i. Where $D_{i,j} = 0$ indicates assignment to a job search assistance and $D_{i,j} = 1$ indicates assignment independent job search.

Let $Y_i^* \in \{0, 1\}$ be the potential employment of jobseeker i, indicating whether they could become employed on their own within 6 months. Y_i^* is only observable when the jobseeker is assigned to independent job search. Let Y_i be the observed employment, such that $Y_i = Y_i^*$ if $D_{ij} = 1$ and $Y_i = 0$ if $D_{ij} = 0$.

For each jobseeker there is a vector of characteristics $X_i \in \mathcal{X}$ observable to both the algorithm and the caseworker and a vector of characteristics $V_{i,j} \in \mathcal{V}$ only observable to the caseworker. I assume that a subset of $V_{i,j}$ is useful for predicting Y_i^*, however a subset of $V_{i,j}$ is not useful, but will be taken into account by caseworkers in solving the assignment problem anyways.

The algorithm predicts the probability of a jobseeker finding a job on their own within six months, i.e. $a(X_i) \equiv Pr(Y_i^* = 1|X_i)$. And will recommend independent job search for all jobseekers with a probability, or profiling score, above some threshold s_a which is determined by a policymaker. The algorithmic recommendation will thus follow the decision rule:

$$A_i = 1[a(X_i) \geq s_a] \tag{1}$$

Assume that caseworkers solve a similar prediction problem taking the algorithmic recommendation A_i, the public information X_i and the private information V_{ij} into account such that $h_j(A_i, X_i, V_{ij}) \equiv Pr(Y_i^* = 1|A_i, X_i, V_{ij})$. Note that caseworkers do not observe the profiling score, but only the assignment recommendation. Each caseworker has an individual evaluation function $h_j(.)$ and an individual assignment thresholds s_j. The caseworkers will follow the decision rule:

$$D_{ij} = 1[h_j(A_i, X_i, V_{i,j}) \geq s_j] \tag{2}$$

The caseworker can deviate from the algorithmic recommendation due to (1) a difference in the threshold s_j which, for a given draw of jobseekers, will determine the assignment rate and can be interpreted as the preference for assigning jobseekers to independent job search.

A caseworker with a strong preference in favor of job search assistance will pick a high threshold s_j, resulting in few jobseekers assigned to independent job search and thus a low assignment rate. (2) A deviation can also be due to a different prediction of reemployment probability, either by exploiting private information not accessible to the algorithm or by evaluating the information differently.

The caseworker assignment rate is defined as the propensity to assign jobseekers to independent job search:

$$S_j = E[D_{ij} = 1] \tag{3}$$

And the caseworker conditional employment rate is defined as:

$$M_j = E[Y_i = 1 | D_{ij} = 1] \tag{4}$$

The algorithm's conditional employment rate at the assignment rate of caseworker j is:

$$M_{a(j)} = E[Y_i^* = 1 | D_{a(j)} = 1] \tag{5}$$

$M_{a(j)}$ will differ from M_j if the caseworker disagrees with the algorithmic prediction or ranking of jobseekers.

The impact of caseworker discretion on the prediction accuracy, at each caseworker's assignment rate, will be measured as:

$$\Delta M_{j,a(j)} = M_j - M_{a(j)} \tag{6}$$

Where $\Delta M_{j,a(j)} > 0$ implies that a caseworker performs better than the algorithm and $\Delta M_{j,a(j)} < 0$ worse than the algorithm.

At a given assignment rate, a difference in prediction, and thus a deviation, will be equivalent to a different ranking of jobseekers in terms of reemployment probabilities. If the caseworker deviations leads to a more accurate ranking of jobseekers, that will in expectation lead to a higher employment rate conditional on being assigned to independent job search.

Thus, caseworker discretion over the algorithmic recommendation lead to more accurate assignment decisions (given the policymakers objective to focus resources on jobseekers with the weakest reemployment probability) if caseworkers obtain a higher conditional employment rate than the algorithm at the caseworker's assignment rate.

3.2 Selective labels problem

Note that the employment potential of jobseeker i, Y_i^*, is only observable when the jobseeker is indeed assigned to independent job search. This selection of outcomes, commonly referred to as the selective labels problem (Kleinberg et al., 2018), is the main complicating factor in estimating the algorithmic counterfactual employment rate and thus the causal impact of caseworker deviation from the algorithmic recommendation. The empirical strategy to deal with this missing data problem is described in section 4.

To make this point clear, I will derive the selection bias when comparing the employment rate of a caseworker j to that of the algorithmic counterfactual at a given cutoff s_a, using the employment outcomes we can observe in the data.

$$\begin{aligned} & E[Y_i^*|D_{ij}=1] - E[Y_i^*|D_i=1, a(X_i) \geq s_a] \\ & = E[Y_i^*|D_{ij}=1] + E[Y_i^*|a(X_i) \geq s_a] - E[Y_i^*|a(X_i) \geq s_a] - E[Y_i^*|D_i=1, a(X_i) \geq s_a] \\ & = \underbrace{M_j - M_a}_{\text{True diff in conditional employment rate}} + \underbrace{E[Y_i^*|a(X_i) \geq s_a] - E[Y_i^*|D_i=1, a(X_i) \geq s_a]}_{\text{Selection bias}} \end{aligned} \tag{7}$$

The selection bias would be equal to zero only if the caseworkers' assignment decision is uncorrelated with the true employment potential, Y_i^*, i.e. if the caseworker assignment decisions were as good as random. An approach relying on a selection-on-observables assumption would be equally problematic, as it would assume that caseworkers are not using any private information to make the assignment decision. While the whole point of this paper is to understand precisely if and how caseworkers use private information to improve on the algorithmic recommendation. In the context of this study, a zero selection bias could only be obtained in an experiment, by randomly overriding the deviation decision of the caseworker.

The selective labels problem introduces selection bias whenever the algorithmic counterfactual conditional employment rates are measured using the directly observable employment rates in the data. Thus there would be selection bias also in other potential specifications, such as directly regressing caseworker deviation rates on employment rate.

4 Empirical strategy

In order to study the causal effect of caseworker deviation, I need to estimate the jobseeker conditional employment outcomes under the algorithmic counterfactual assignment. Using quasi random assignment of jobseekers to caseworkers, I estimate the algorithmic employ-

ment rate by exploiting variation in caseworkers assignment rate.

The main complicating factor for estimating the employment rate under the algorithmic counterfactual is a missing data problem - there will be selection in whose outcomes are observable. Specifically, a jobseeker's employment outcome under independent job search is not observable if the jobseekers was assigned to job search assistance. This missing data problem is commonly referred to as the *selective labels problem* in the literature, see e.g. Kleinberg et al. (2018), and has been a major obstacle for estimating the causal impact of algorithms using observational data.

To address the selective labels problem, I will follow the extrapolation approach suggested by Angelova, Dobbie, and Yang (2023). Angelova, Dobbie, and Yang (2023) study human discretion over an algorithm in the context of bail hearings in the US. The identification strategy relies on random assignment of jobseekers to caseworkers in order to ensure unbiased estimates, and differences in caseworkers assignment rates to get exogenous variation in the assignment of jobseekers. Note that since almost all caseworkers have high assignment rates, and 15 % have assignment rates equal to 1, the extrapolation should yield similar result to worst case bounding.

In order to compare performance between the algorithm and caseworkers I will estimate the average jobseeker employment potential for each assignment rate, S, such that all caseworkers assignment rates are spanned.

Let $M_j = E[Y_i^*|D_{ij} = 1]$, be the conditional employment rate for caseworker j. And let $S_j = E[D_{ij}]$ be the caseworker's specific assignment rate. In order to compare performance, I will construct an algorithmic counterfactual employment rate for each caseworker at the caseworker's specific assignment rate. Note that for the algorithm, each assignment rate S is associated with a threshold of the profiling score s. Let the threshold needed for the algorithm to have the same assignment rate as caseworker j be $s_{a(j)} = \{s : Pr(a(X_i) \geq s) = Pr(h_j(.) \geq s_j)\}$

The counterfactual algorithmic employment rate at the assignment rate of caseworker j will be:

$$M_{a(j)} = E[Y_i^*|D_{i,a(j)} = 1] \tag{8}$$

4.1 Estimating the caseworkers parameters

The caseworker-specific assignment rates and employment rates can be estimated using the following specification on the full sample.

$$D_i = S_j + \phi_o \times \gamma_t + u_{iot} \tag{9}$$

$$Y_i = M_j + \phi_o \times \gamma_t + v_{iot} \tag{10}$$

Where D_i and Y_i are indicators for assignment to independent job search and employment for jobseeker i, respectively. S_j is the estimated assignment rate and M_j is the estimated conditional employment rate of caseworker j. The model also includes interacted office, ϕ_o, and and quarter, γ_t, fixed effects. There are 10 offices. Standard errors are constructed using the bootstrapping.

The caseworker employment rate will be estimated with sampling error, due to a finite number of observations per caseworker. The sampling error will inflate the variance in the caseworker employment rate, which could bias the comparison to the algorithmic counterfactual. To correct for the sampling error I use empirical Bayes shrinkage, following Kane and Staiger (2008). By applying the shrinkage procedure, the point estimates of each caseworker are shrunk towards the mean of the estimated fixed effects.

4.2 Estimating the algorithmic counterfactual

4.2.1 Intuition

In order to provide intuition for the estimation of the algorithmic counterfactual employment rate, consider an idealized setting where there is quasi-random allocation of jobseekers to caseworkers and where there is one caseworker, caseworker j^*, who assigns all jobseekers at or above a given threshold s_a to independent search. Such behavior is consistent with (i) having a strong preference for assigning jobseekers to independent job search, such that their assignment rate is close to one, or (ii) with the caseworker being totally compliant with the algorithm at the threshold s_a.

Regardless of whether it is because of preference on assignment or compliance with the algorithm, the assignment rate at the threshold s_a will be equal to one. With random assignment of jobseekers to caseworkers, the employment rate among that subsample of jobseekers should be close to the average employment rate for the subset of jobseekers with profiling scores at or above the threshold.

In the absence of a caseworker with assignment rate equal to one, the employment potential can be estimated using extrapolation to the cut-off. In order to ensure that the extrapolation provide a consistent estimate of the average employment potential in each sample, there need to be sufficiently many caseworkers with assignment rates close to 100 % in each sub-sample. See Figure 8 for confirmation that this is indeed the case.

4.2.2 Implementation

The implementation of the empirical strategy can be summarized in the following steps:

1. **Restrict the sample to jobseekers with scores at or above the given threshold,** $a(x_i) \geq s$
   For the algorithm, we know that each assignment threshold, s, is associated with some algorithmic assignment rate, S. Thus, etimating the algorithmic counterfactual employment rate at some assignment rate, S, is equivalent to doing so for the subset of jobseekers with score at or above the associated assignment threshold.

2. **Estimate the caseworker level assignment rates and employment rates within the sub-sample.** Controlling for office and time fixed effects.

$$D_i = S_j + \phi_o \times \gamma_t + u_{iot}$$

$$Y_i = M_j + \phi_o \times \gamma_t + v_{iot}$$

3. **Estimate the average employment potential in the sub-sample**
   Regress the estimated caseworker employment rates on assignment rates. The average employment potential within the sub-sample of jobseekers with $a(x_i) \geq s_a$ is equivalent to the employment rate of a caseworker with an assignment-rate of 100 % within the sub-sample (due to random assignment of jobseekers to caseworkers). In the absence of such a caseworker, we need to estimate the employment rate using extrapolation.

   Center the assignment rate estimates at $S_j = 1$, i.e. by defining a new variable $\hat{S}_j^C = \hat{S}_j - 1$. Then run the regression:

$$\hat{M}_j = \beta_0 + \beta_1 \hat{S}_j^C + e_i \tag{11}$$

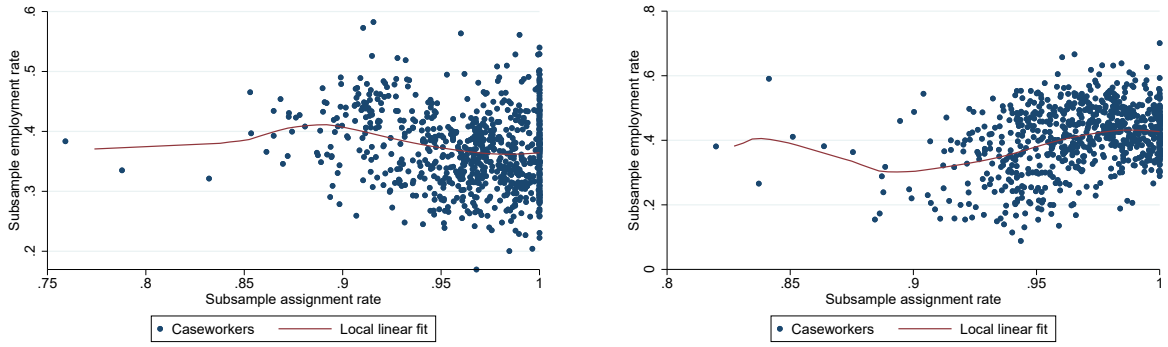
   Due to centering around $\hat{S}_j = 1$, β_0 is the estimated employment rate at an assignment rate of 100 % within the sub-sample, which provides the estimate of the threshold-specific average employment rate in the sub-sample of jobseekers with $a(x_i) \geq s$. That is, β_0 provides the estimate of $M_{a(j)}$.

4. **For robustness:** do this procedure using (i) linear extrapolation (OLS) and (ii) local linear extrapolations.

Repeat the procedure for many thresholds such that the whole span of caseworker assignment rates is covered. The estimate of the population average potential employment can be

obtained by using the whole sample of jobseekers and extrapolate to the vertical intercept on the right hand side, i.e. at $S_j = 1$.

Figures below show the extrapolation to assignment rate equal to one for two subsamples.



(a) Assignment rate S=100%, a(X_i) ≥ 0

(b) Assignment rate S=63 %, a(X_i) ≥ 0.45

Figure 6: Subsample assignment rates and conditional employment rates.

Note: Estimation using local linear regression. Caseworker assignment rate and conditional employment rates estimated using office-quarter fixed effects. Includes caseworkers with at least 75 profilings per year.

4.3 Identifying assumptions

The identifying assumptions needed for estimating the causal impact of caseworker discretion over an algorithmic recommendation, and interpret the results are (i) random assignment of jobseekers to caseworkers, (ii) first stage - caseworker assignment rate should impact the probability of being assigned to independent job search, (iii) caseworkers with high assignment rates within each subsample of interest, (iv) exclusion restriction, and (v) caseworker objectives are well-specified.

A major advantage of the empirical approach is that no monotonicity assumption is needed for the first stage. That is, it does not require that a jobseeker who is assigned to independent job search by a caseworker with a low assignment rate, is always assigned to independent job search by a caseworker with a higher assignment rate. In other words, it would require that caseworkers have identical ranking of jobseekers. A monotonicity assumption would be particularly unlikely to be satisfied in this context, and would in fact not be inconsistent with the conceptual framework which has as a central idea that private information as well as the evaluation function vary across caseworkers. The difference in evaluation of the information at hand can be due to either ability to correctly interpret the information or due to preferences.

It is enough that the assignment rate impacts the probability that an individual jobseekers is assigned to independent job search, that is that the "first stage" is significant.

Note also that the method is agnostic with regards to the effectiveness of job search assistance, as the employment outcomes will be based solely on jobseekers assigned to independent job search.

4.3.1 First stage and caseworker assignment rate

Table 2 shows the impact of the caseworker leave-out assignment rate on the probability that a jobseeker is assigned to independent job search. It shows that being assigned to a caseworker with a higher assignment rate, significantly increases the probability that a single case gets allocated to independent job search.

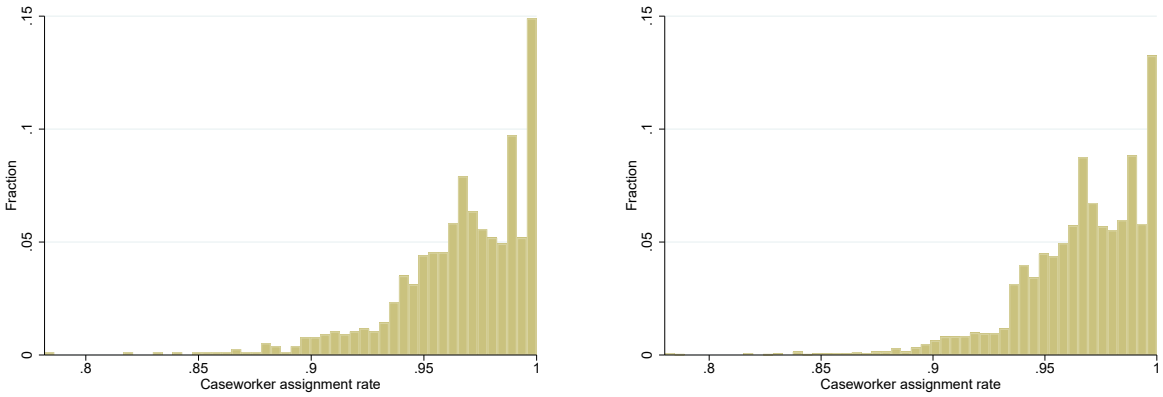
Table 2: First stage

	Probability independent jobsearch
CW assignment rate	0.434***
	(0.040)
N	136,392
Mean assignment rate	0.97
Number of caseworkers	772
Number of jobseekers	121,968

Note: Standard errors are clustered at caseworker level and reported in parentheses. ***, **, * indicate significance at 1%, 5% and 10% respectively. Regressions include office-quarter fixed effects. Dependent variable is the leave-out caseworker assignment rate. Only includes caseworkers with at least 75 profilings per year.

Figure 7 shows that there is significant variation in the caseworkers assignment rate, ranging from 0.79 to 1, with a mean of 0.97. Given the sample selection, e.g. that we are zooming in on jobseekers who are getting service at the central offices, it is not surprising that the assignment rates are relatively high.

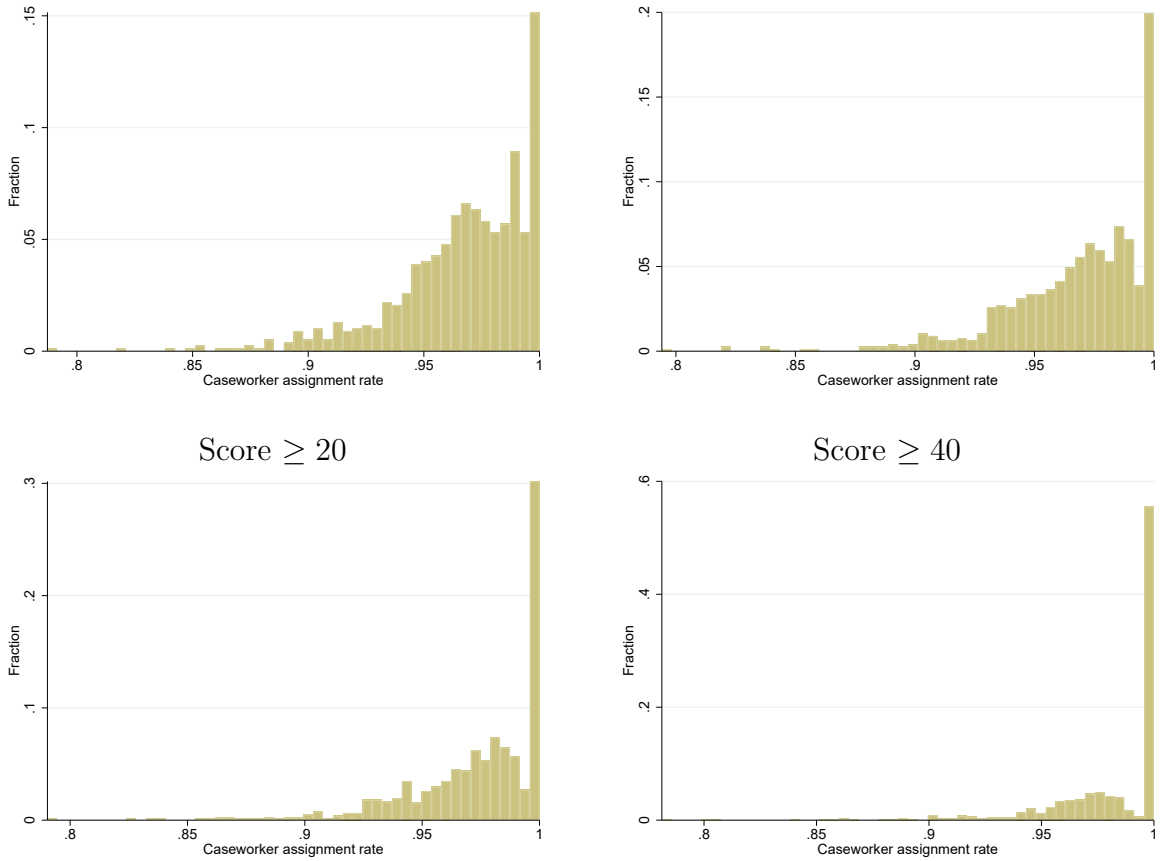
Figure 8 shows the distribution of assignment rates at caseworker level for four subsamples. Each subsample consists of all jobseekers with profiling scores above the given cut-off and the assignment rates are calculated within the specific subsample. A higher profiling score indicates a higher reemployment probability. The figure illustrates how the distribution of assignment rates shift to the right as the cut-off increases and the subsample gets more and more restrictive. For example, more than 30 % of caseworkers assigns all jobseekers with scores above 60 to independent job search.



Caseworker level Case level

Figure 7: Caseworker assignment rate

Note: Includes caseworkers with at least 75 profilings per year.



Score $\geq$ 20 Score $\geq$ 40

Score $\geq$ 50 Score $\geq$ 60

Figure 8: Caseworker assignment rates within subsamples

Note: Includes caseworkers with at least 75 profilings per year.

4.3.2 Random assignment of jobseekers to caseworkers

The key identifying assumption for the paper is random assignment of jobseekers to caseworkers. The institutional design of assignment of jobseekers at the Swedish PES central offices suggests that the assumption of random assignment holds. Table 3, column 1, shows the result of single regressions of caseworker leave-out assignment rate [8] on predetermined jobseeker characteristics. Column 2 shows the result of single regressions of the profiling score on predetermined jobseeker characteristics. The results show that while jobseeker characteristics are significantly correlated with the profiling score, they are uncorrelated with the leave-one-out caseworker assignment rate. The F-test of joint significance of jobseeker characteristics confirms the findings of the single regressions. Suggesting that there is no sorting between jobseekers and caseworkers and that the assumption of random assignment holds.

[8] $S_{ij}^{leave-out} = \frac{\sum_{i=1}^{N} D_{ij} - D_{ij}}{N-1}$

Table 3: Test random assignment of jobseekers to caseworkers

	Assignment rate	Profiling score
Male	0.000	0.027***
	(0.000)	(0.001)
Age	-0.000	-0.002***
	(0.000)	(0.000)
Historical functional disability	-0.001	-0.133***
	(0.000)	(0.002)
Immigrant	-0.001**	-0.064***
	(0.000)	(0.001)
High school	0.001	0.015***
	(0.001)	(0.003)
Tertiary education	0.000	0.017***
	(0.000)	(0.001)
Profiling score	0.002*	1.000***
	(0.001)	(0.000)
F-test	1	2,294***
N	136,392	136,392
N caseworkers	772	772
Mean a-rate and score	0.97	0.51

Note: Standard errors are clustered at caseworker level and reported in parentheses. ***, **, * indicate significance at 1%, 5% and 10% respectively. Regressions include office-quarter fixed effects. Assignment rate is the leave-out mean. The F-test for joint significance of jobseeker characteristics on the profiling score excludes the profiling score. Only includes caseworkers with at least 75 profilings per year.

5 Results

Table 4 shows the share of caseworkers who obtain a higher conditional employment rate than the algorithm and than a random assignment, holding the assignment rate constant. The results tell us that by deviating from the algorithmic recommendation, the average caseworker obtains the same conditional employment rate as had they not deviated. But there is substantial performance heterogeneity across caseworkers. 52 % of caseworkers perform worse than the pure algorithmic allocation of jobseekers, and 49 % of caseworkers

perform worse than a random allocation of jobseekers. And both high and low performing caseworkers are deviating to the same extent.

Table 4: Caseworker vs algorithmic prediction accuracy

	Share $M_j > M_{a/r}$	Average $\Delta M_{j,a/r}$
Algorithmic counterfactual	0.482	0.003
Random assignment	0.513	0.011
N	772	772

Note: Caseworker vs algorithmic counterfactual, holding the assignment rate constant. Column 1 shows the share of caseworkers with a higher conditional employment rate than the algorithmic counterfactual and random assignment. Column 2 shows the weighted average difference between caseworker conditional employment rate and algorithmic counterfactual and random assignment. Employment outcome is full time employment within six months lasting four months, based on monthly payroll tax data. Results are robust to weighting by caseworker case load. Estimates are generates using office-quarter fixed effects.

Figure 9 illustrates the estimated performance of caseworkers and the algorithm. It shows the conditional employment rate, holding assignment rate fixed. Each point in the scatter plot corresponds to a caseworker. The orange line shows the average conditional employment rate obtain by caseworkers at any given assignment rate, estimated using local linear regression. Similarly, the blue line shows the estimated algorithmic counterfactual conditional employment rate at a given assignment rate, using local linear regression.

The dashed line corresponds to the conditional employment rate under a random allocation of jobseekers, which is simply constant and equal to the estimated population employment potential. The estimated population employment potential is given by the estimated algorithmic counterfactual at assignment rate equal to one. By construction the caseworker average, the algorithmic counterfactual and the random allocation will reach the same expected employment rate at an assignment rate of one.

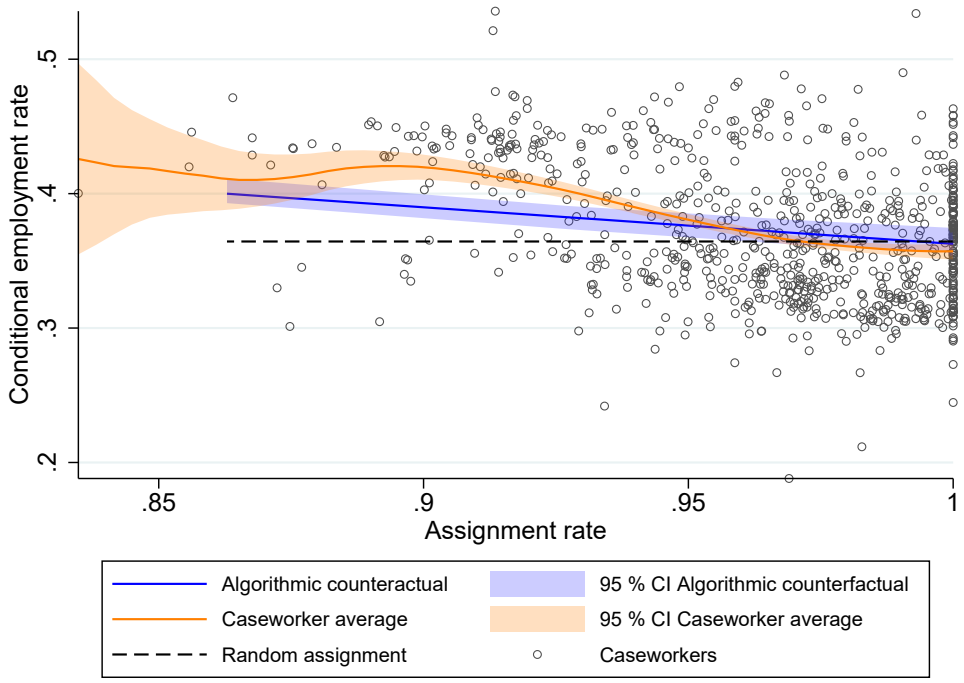


Figure 9: Full time employment within six months lasting four months

Note: Employment outcomes are based on monthly payroll tax data. The fitted lines are estimated using local linear regression. Algorithmic counterfactual estimated using local linear regression within each subsample. Estimates are generated using office-quarter fixed effects. Caseworker conditional employment rate is estimated using empirical Bayes shrinkage. Includes caseworkers with at least 75 profilings per year.

If a decision maker is good at ranking jobseekers by their reemployment probability, we would expect to see a negative relationship between the conditional employment rate and the assignment rate. The intuition is that a decision maker with a low assignment rate, who is more selective, should assign the jobseekers with the highest reemployment probability to independent job search first. So that their pool of jobseekers, for whom the employment rate is calculated, are more likely to become employed on their own. Notably, there is a negative relationship between conditional employment rate and assignment rate for the algorithm, with a point estimate of the slope from a linear regression of -0.272 (p-value = 0.001). Across caseworkers there is a negative relationship between conditional employment rate and assignment rate, with a point estimate from a linear regression of -0.601 (p-value < 0.000). Which implies that caseworkers are on average better than the algorithm at predicting the reemployment probability of jobseekers.

However, Figure 9 also illustrates one important margin of performance heterogeneity. For assignment rates above 95 % both the algorithm and caseworkers are unable to rank jobseekers based on their reemployment proabability, and obtain a conditional employment

rate not much better than a random assignment. Meanwhile caseworkers are outperforming the algorithm at lower assignment rates.

To inform human oversight policy, it is crucial to gain deeper insights into the mechanisms driving performance heterogeneity among caseworkers. The following subsections will examine this issue in terms of (i) reasons for deviation, (ii) caseworker characteristics and use of information, and (iii) jobseeker characteristics. Additionally, I will explore the importance of caseworker objectives.

5.1 Reasons for deviation

Caseworkers are required to report their reason for deviating, typically through free-text responses. Reporting is only required when the algorithm recommended job search assistance but the caseworker deviated to independent job search. Figure 10 shows the entered reasons for deviations. Notably the most common reason for deviating is that the jobseeker (i) is already working to some degree, or (ii) will start a job within the next few months. It is puzzling that despite this being the most common reason for deviating, the average caseworker is not significantly improving on the algorithmic prediction. I will further investigate how caseworker prediction performance vary with the reason behind deviating.

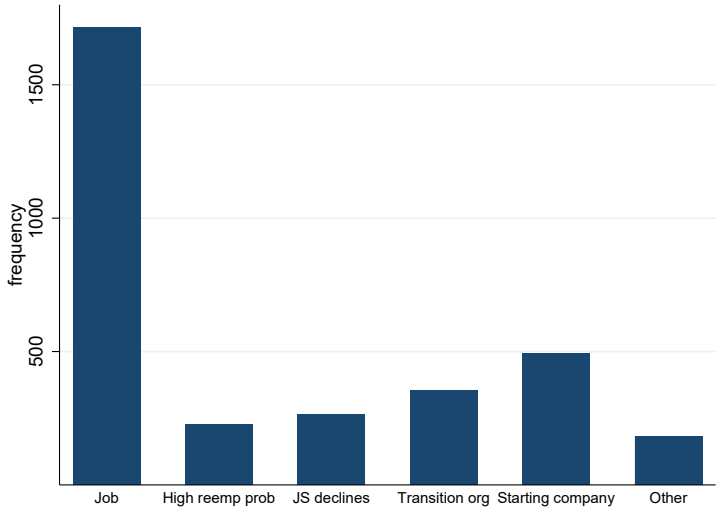


Figure 10: Reasons for Caseworker Deviation

Note: Caseworker reasons for deviating, when the algorithm recommended job search assistance and the caseworker instead assigned to independent job search. Total number of deviations are from job search assistance to independent search is 3,228. 'Job' indicates that the jobseeker (i) is already working to some degree, or (ii) will start a job within the next few months. 'High reemp prob' indicates high reemployment probability assessed by the caseworker. 'JS declines' means that the jobseeker has declines the offer with a sanction to UI benefits as a consequence. 'Transition org' means that the jobseeker is already receiving help from a transition organization financed by the employer and union. 'Starting company' means that the jobseeker are starting a company.

Caseworkers are neither required nor equipped to verify whether a jobseeker claiming to work part-time or start a new job soon is telling the truth. In order to better understand this I have plotted the evaluation of the median monthly earnings for the jobseekers. The median jobseeker in the group of jobseekers who claim to work part-time or start a new job shortly earn significantly more compared to the other groups of jobseekers. They have median monthly earnings of SEK 20,000 ($\approx 1,860$ USD) one month after the profiling. Implying that jobseekers are on average telling the truth. Jobseekers assigned to job search assistance have the lowest monthly earnings, which could reflect the lower underlying reemployment probability as well as a locking-in effects, i.e. that starting job search assistance delays reentry to the labor market.

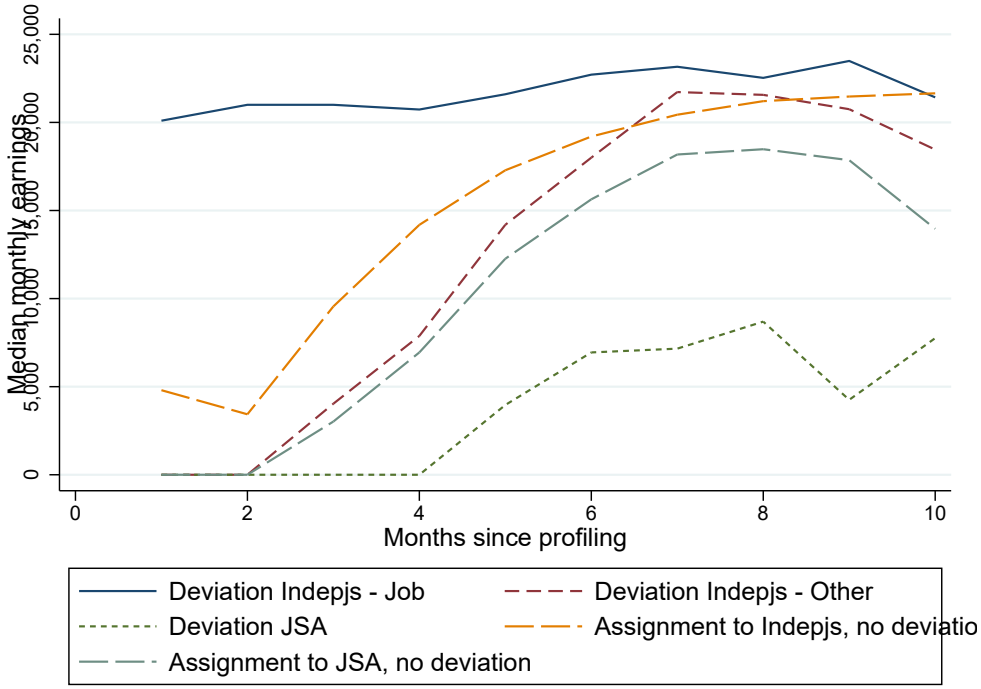


Figure 11: Median monthly earnings since profiling

Note: Monthly earnings in SEK per jobseeker based on payroll tax data.

References

Angelova, V., W. S. Dobbie, and C. Yang (2023): "Algoriothmic recommendations and human discretion," *NBER working paper*, September 2023, 1–87.

Arbetsförmedlingen (2021): *Kundval Rusta och Matcha, Uppföljning och lärdomar efter ett år med en ny matchningstjänst*, Arbetsförmedlingen.

——— (2023): *Arbetsförmedlingens nya modell för bedömning av arbetssökandes stödbehov*, Arbetsförmedlingen.

Arni, P., G. J. van den Berg, and R. Lalive (2022): "Treatment versus regime effects of carrots and sticks," *Journal of Business & Economic Statistics*, 40, 111–127.

Arnold, D., W. Dobbie, and P. Hull (2022): "Measuring racial discrimination in bail decisions," Tech. rep.

Behncke, S., M. Frölich, and M. Lechner (2010): "A caseworker like me–Does the similarity between the unemployed and their caseworkers increase job placements?" *The Economic Journal*, 120, 1430–1459.

Cederlöf, J., M. Söderström, and J. Vikström (2021): "What makes a good caseworker?" Tech. rep., Working Paper 2021: 9, IFAU, Uppsala.

Chan, D. C., M. Gentzkow, and C. Yu (2022): "Selection with variation in diagnostic skill: Evidence from radiologists," *The Quarterly Journal of Economics*, 137, 729–783.

Cowgill, B. (2020): "Bias and productivity in humans and algorithms: Theory and evidence from resume screening," *Columbia Business School, Columbia University*, 29.

Dawes, R. M. (1979): "The robust beauty of improper linear models in decision making." *American psychologist*, 34, 571.

Dawes, R. M., D. Faust, and P. E. Meehl (1989): "Clinical versus actuarial judgment," *Science*, 243, 1668–1674.

Egebark, J., L. Laun, L. Liljeberg, M. Rödin, M. Söderström, E. Videnord, and J. Vikström (2024): "En effektutvärdering av arbetsförmedling med fristående leverantörer," Tech. rep., IFAU-Institute for Evaluation of Labour Market and Education Policy.

Ernst, S., A. I. Mueller, and J. Spinnewijn (2024): "Risk score for long-term unemployment and the assignment to job search counceling," *Working paper*.

Grove, W. M., D. H. Zald, B. S. Lebow, B. E. Snitz, and C. Nelson (2000): "Clinical versus mechanical prediction: a meta-analysis." *Psychological assessment*, 12, 19.

Hoffman, M., L. B. Kahn, and D. Li (2018): "Discretion in hiring," *The Quarterly Journal of Economics*, 133, 765–800.

Kane, T. J. and D. O. Staiger (2008): "Estimating teacher impacts on student achievement: An experimental evaluation," Tech. rep., National Bureau of Economic Research.

Kleinberg, J., H. Lakkaraju, J. Leskovec, J. Ludwig, and S. Mullainathan (2018): "Human decisions and machine predictions," *The quarterly journal of economics*, 133, 237–293.

Lechner, M. and J. Smith (2007): "What is the value added by caseworkers?" *Labour economics*, 14, 135–151.

Ludwig, J. and S. Mullainathan (2021): "Fragile Algorithms and Fallible Decision-Makers: Lessons from the Justice System," *Journal of Economic Perspectives*, 35, 71–96.

Mullainathan, S. and Z. Obermeyer (2022): "Diagnosing physician error: A machine learning approach to low-value health care," *The Quarterly Journal of Economics*, 137, 679–727.

OECD (2019): "Statistical profiling in public employment services," .

Ohlin, L. E. and O. D. Duncan (1949): "The efficiency of prediction in criminology," *American Journal of Sociology*, 54, 441–452.

Rambachan, A. (2024): "Identifying prediction mistakes in observational data," *The Quarterly Journal of Economics*, qjae013.

Rambachan, A., A. Coston, and E. Kennedy (2024): "Robust Design and Evaluation of Predictive Algorithms under Unobserved Confounding," *Working paper*.

van den Berg, G. J., M. Kunaschk, J. Lang, G. Stephan, and A. Uhlendorff (2023): "Predicting re-employment:machine learning versus assessments by unemployed workers and by their caseworkers," *IFAU working paper*, 1–58.

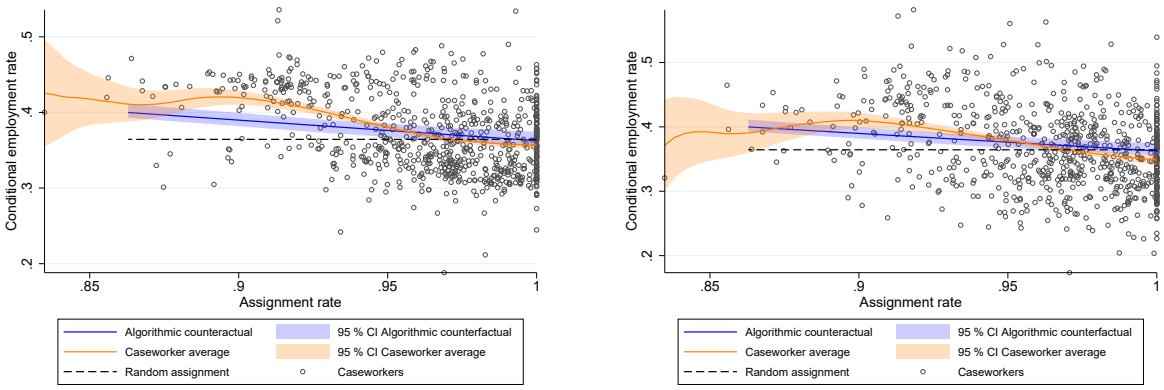
A Appendix A

Table A1: Input variables algorithm

Date of registration PES
Country of birth
Disability
Number of PES programs or interventions 10 years
Number of days in PES programs or interventions 10 years
Number of days registered at the PES, current or the numebr of days registered at the latest registratio
Number of days since the last registration three
Age
Municipality
Zip code
Search category
Target occupation
Education in target occupation
Experience in target occupation
Searching for work to the extent of (full time, part time)
Level of education
Field of education
Gender
Applying for jobs abroad
Applying for unemployment insurance
Share of unemployed within the zip code
Median income zip code
Share with secondary education within zip code
Share of households types within zip code
Share with non-European citizenship within zip code.

B Appendix B

B.1 Results without Empirical Bayes shrinkage



With EBS Without EBS

Figure B1: Results with and without empirical Bayes shrinkage

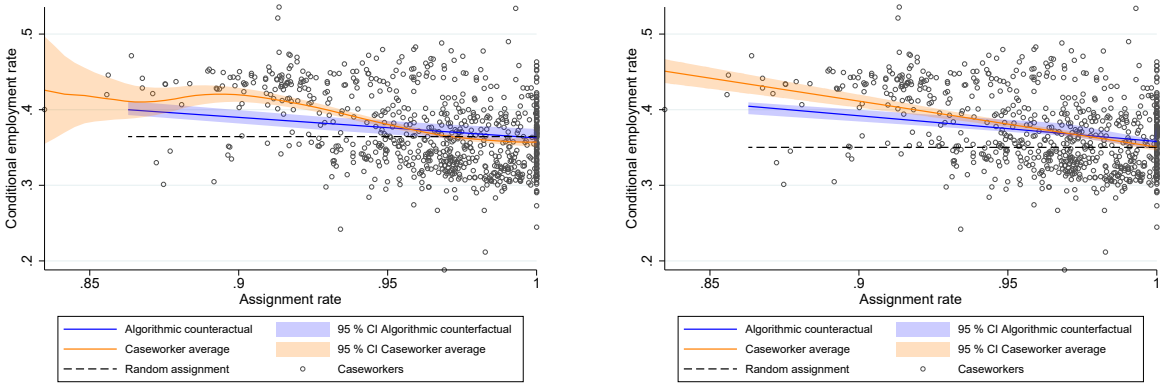
Note: Caseworker conditional employment rates estimated with and without Empirical Bayes Shrinkage. Employment outcomes are based on monthly payroll tax data. The fitted lines are estimated using local linear regression. Estimates are generated using office-quarter fixed effects. Includes caseworkers with at least 75 profilings per year.

Table B1: Caseworker vs algorithmic prediction accuracy - No EBS

	Share $M_j > M_{a/r}$		Average $\Delta M_{j,a/r}$	
	With EBS	Without EBS	With EBS	Without EBS
Algorithmic counterfactual	0.482	0.478	0.003	-0.003
Random assignment	0.513	0.509	0.011	0.006
N	772	772	772	772

Note: Caseworker employment rate vs algorithmic counterfactual employment rate, holding the assignment rate constant. Columns 1 -2 show the share of caseworkers with a higher conditional employment rate than the algorithmic counterfactual and random assignment. Column 3-4 show the weighted average difference between caseworker conditional employment rate and algorithmic counterfactual and random assignment. Full time employment within six months lasting four months. Employment outcomes are based on monthly payroll tax data. Estimates are generates using office-quarter fixed effects. Includes caseworkers with at least 75 profilings per year.

B.2 Robustness: local linear regression



Local linear regression Linear regression

Figure B2: Results using linear regression

Note: Algorithmic counterfactual and across caseworker relationship estimates using local linear regression and linear regression. Employment outcomes are based on monthly payroll tax data. Estimates are generated using office-quarter fixed effects. Includes caseworkers with at least 75 profilings per year.

C Appendix C

C.1 Empirical Strategy

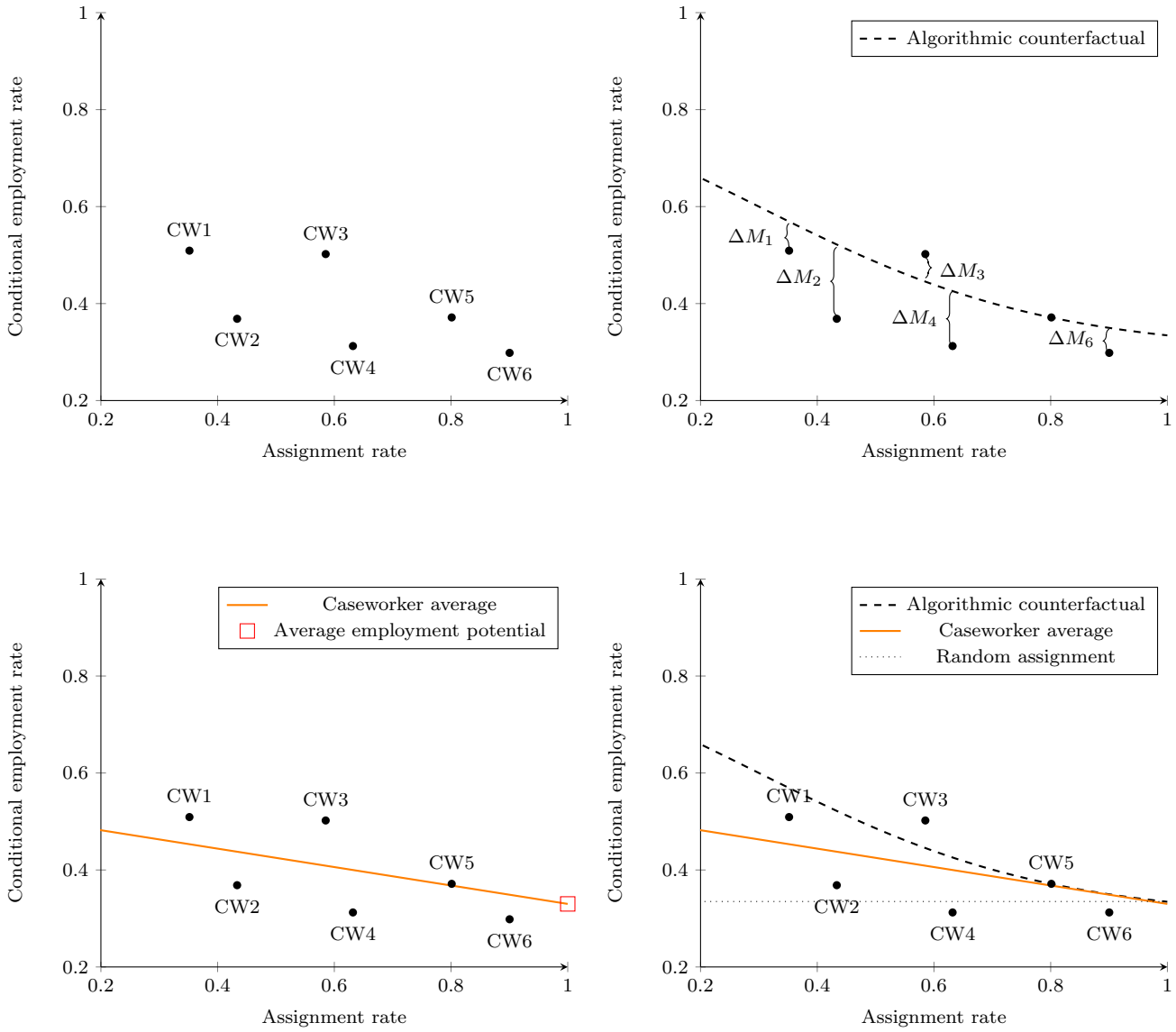


Figure C1: Empirical Strategy step by step

Note: Conditional employment rate is defined as $M_j = E[Y_i = 1|D_{ij} = 1$. Assignment rate is defined as $S_j = E[D_{ij} = 1]$.