A policy learning approach to fair job candidate recommendations

Achim Ahrens*

November 15, 2024

— Early draft. Please do not circulate. —

Abstract

Recommender systems are increasingly popular in public policy. However, there is a that the naïve adoption of recommender systems can have unintended consequences. This study designs a recommender system that optimizes job candidate suggestions under fairness and congestion constraints. The recommender builds on a policy learning framework and is trained on rich vacancy-job seeker data from Swiss employment centers. The results highlight two risks associated with recommender systems for job matching. First, an unconstrained recommender system may violate specific notions of fairness, even if the training data outcomes are consistent with such notions of fairness. Second, recommender systems may aggravate the concentration of recommendations across vacancies, which likely undermines the efficacy of the recommender system. Preliminary results suggest that both concerns can be addressed in the proposed policy learning framework.

Keywords: Fairness, discrimination, candidate recommendations, policy learning.
JEL: J15, J16, J64, D63, C44

*ETH Zürich, Switzerland. *Email:* `achim.ahrens@gess.ethz.ch`.

1 Introduction

Advances in machine learning, the increasing availability of high-performance computing power, and the emergence of rich data sets have led to the popularity of recommender systems for public policy, e.g., for predicting recidivism (Dressel and Farid, 2018), the risk of long-term unemployment (Mueller and Spinnewijn, 2023), or optimizing the geographic allocation of refugees (Bansak et al., 2018). Machine learning tools promise to automate decision-making processes and detect predictive signals otherwise invisible to human decision-makers. However, there is a risk that recommender systems unintendedly discriminate against protected characteristics such as gender or race (Angwin et al., 2016; Barocas, Hardt, and Narayanan, 2023), raising both legal and ethical concerns.

A central domain within economics where the use of algorithmic recommendation systems is considered is the matching of vacancies and job seekers (Kircher, 2022). While linked employer-employee register data or user data from online job platforms provide suitable training data, there are substantial risks to naïvely adopting recommender systems. It is well documented that hiring discrimination in the labor market by age, gender, and race is commonplace (e.g., Neumark, 2018; Hangartner, Kopp, and Siegenthaler, 2021; Lippens, Vermeiren, and Baert, 2023), implying that a recommender system trained on historical data may reinforce discriminatory biases. Even if observed hiring decisions are exclusively driven by productivity-relevant signals, recommender systems may yield differential classification error rates across groups (Zafar et al., 2017).

The literature on fairness and artificial intelligence has proposed a variety of fairness notions, including demographic parity, predictive parity, fairness through awareness (Dwork et al., 2011), counterfactual fairness (Kusner et al., 2018), equalized odds (Hardt et al., 2016; Zafar et al., 2017).[1] Yet, little is known about how imposing notions of fairness affects the effectiveness of job recommendations. In this research, I design an algorithm that recommends job candidates to firms and allows for fairness constraints. I focus on demographic parity (balance in recommendation rates across groups) and predictive parity (balance in expected hirings) with respect to gender, nationality (Swiss vs.

[1] For a review, see Castelnovo et al. (2022) and Barocas, Hardt, and Narayanan (2023).

non-Swiss), and the length of the job seeker's search spell. I empirically illustrate the trade-off between the two notions of fairness as well as their impact on expected hiring. Mapping trade-offs between notions of fairness and efficiency informs the debate on which fairness metrics are most appropriate for designing job recommender systems. Furthermore, I examine the problem of congestion, highlighted in Bied et al. (2022), which refers to the risk that a recommender system concentrates a large share of recommendations on a few candidates, undermining the efficacy of the system.

Building on the literature on statistical decision-making (e.g., Manski, 2004; Athey and Wager, 2021), also referred to as policy learning, I estimate policy rules that decide when a candidate should be recommended for a particular position while allowing for cost, fairness, and concentration constraints. The most general policy rule assigns recommendations based on predicted expected hiring, group membership, and the candidate's rank. In the application, I consider the case of Switzerland, where, for some occupations, caseworkers at regional employment centers are mandated to find suitable job candidates to fill open vacancies. In this setting, hiring is a two-stage process: caseworkers first decide whether to recommend a candidate based on both protected and productivity-relevant signals. In the second stage, employers make a hiring decision.

The analyses reveal that an unconstrained recommender system may fail to comply with demographic parity and predictive parity, even though caseworkers' recommendations are approximately consistent with predictive parity. For example, an unconstrained recommender system is 5.2 ppts. less likely to recommend female job seekers compared to male job seekers, while the gap in caseworkers' recommendation rates amounts to only 1 ppt. The proposed framework allows imposing either demographic parity or predictive parity, but not both due to their inherent incompatibility. Importantly, enforcing fairness does not lead to a meaningful drop in expected hiring. A drawback of recommender systems seems to be that they can aggravate the concentration of recommendations across vacancies. I show that this issue can be addressed by imposing additional Gini-based concentration constraints at the cost of a reduction in the recommendations' expected hiring chances.

The study contributes to the recent literature exploring the feasibility and benefits of machine-learning methods for job recommendations (e.g., Li, Raymond, and Bergman, 2020; Bied et al., 2022; Behaghel et al., 2024). In particular, Barbanchon, Hensvik, and Rathelot (2023) deploy a collaborative filtering system on a Swedish online job platform and test its performance in a randomized experiment. The setting in this study differs in that I focus on caseworkers' recommendations at regional employment centers, which allows me to leverage rich jobseeker characteristics and observed hiring decisions. Li, Raymond, and Bergman (2020) consider the trade-off in exploitation (recommending groups that have historically performed well) and exploration (considering unknown groups to learn about their performance). They show that putting more emphasis on exploration can increase both performance and diversity. Bied et al. (2022) show that the use of recommender systems for French Public Employment services can increase congestion, reflected in the same job ad frequently being recommended to many candidates, et vice versa. Most similar to this study are Rus et al. (2022) and Bied et al. (2023), who consider an adversarial approach to decorrelate predictive features from gender. I differ from these studies in that I address both congestion and fairness aspects in a unified policy learning framework that leverages generic supervised machine learners.

The article lies at the intersection of the literature on algorithmic fairness and policy learning. The field of algorithmic fairness originated in computer science (Dwork et al., 2011; Kusner et al., 2018; Hardt et al., 2016; Zafar et al., 2017) but has more recently received attention in economics (Kleinberg et al., 2018; Ludwig and Mullainathan, 2021; Arnold, Dobbie, and Hull, 2024). The econometric field of statistical decision-making goes back to Wald (1950) and Savage (1951) and has been reinvigorated by Manski (2004). The aim of the field is to estimate optimal rules that allocate treatments to individuals based on observed characteristics. While earlier work assumed samples generated by randomized experiments, the recent literature allows for non-random treatment assignment in the training data (Manski, 2007; Athey and Wager, 2021). This study adds to the few theoretical studies studying fairness in the context of estimating statistical decision rules (Viviano and Bradic, 2022; Kim and Zubizarreta, 2023).

I proceed as follows. The next section describes the institutional background and the data. Section 3 estimates observed and unexplained gaps in hirings with respect to the relevant characteristics. Section 4 describes the estimation methodology. The results are discussed in Section 5. Section 6 concludes.

2 Institutional background and data

2.1 Background

Following the 2014 referendum on the popular initiative "Against Mass Immigration," the Swiss parliament adopted a set of measures intended to increase the hiring chances of the domestic workforce relative to jobseekers from abroad. As part of these measures, employers must register vacancies with regional employment centers (*Regionale Arbeitsvermittlungszentren*; hereafter RAVs) if the occupational unemployment rate is above 5% (above 8% in 2018–2019). To improve hiring chances among domestic jobseekers, affected vacancies are only visible to jobseekers living in Switzerland for the first five working days. Caseworkers at the RAVs have three working days to search for suitable candidates and may express candidate recommendations to employers.[2] If caseworkers are not able to find a suitable candidate, they have to notify the employer. In case of a recommendation, employers are asked to provide feedback on whether the candidate recommendation leads to a hiring decision (88.5% response rate). While caseworkers may also recommend candidates outside of the designated occupation, 83.2% recommendations are filed in occupations under the notification obligations. In this study, I focus on vacancies that are subject to the notification obligation.

2.2 Data

The data is drawn from AVAM, the software system for job placements and labor market statistics used by the RAVs. The data includes the universe of vacancies and job seekers

[2]The set of measures is evaluated in Ahrens et al. (2021), Sheldon and Wunsch (2021), Bamert et al. (2021), and Braun-Dubler et al. (2021).

registered with the RAVs and covers the period from January 2021 to September 2023. In total, there are 348 663 unique vacancies and 793 117 individuals (over 1 038 705 job search spells) in the sample. The total number of recommendations is 1 060 925 and 3.1% of these recommendations result in a hiring.

AVAM records information on each vacancy's associated occupation (ISCO), industry, workload, preferred age, and gender, requirements for private car, driving licenses, special work schedules (e.g., night shift), language requirements, required work experience (in years), as well as requirements for diploma or qualifications. Appendix Table A.1–A.2 provide summary statistics. AVAM also includes the counterpart of these characteristics on the side of the job seekers, including each job seeker's qualification and experience recorded at the occupation level. In addition, AVAM has information on each jobseeker's nationality, mother tongue, preferred region, and work schedule (see Appendix Tables A.3–A.4).

2.3 Sampling strategy

A major computation challenge in fitting a recommender system to the data is that considering all possible combinations of vacancies and jobseekers at a given time leads to a computationally intractable sample size. In a given month, there are between 148 176 and 256 163 registered jobseekers and between 4735 and 13 851 available vacancies. Even after filtering on vacancy-jobseeker pairs that align with regional preferences, the sample size renders training machine learning algorithms practically infeasible. Another issue is that, with 26 941 occurrences per month, recommendations are relatively rare events compared to all possible vacancy-jobseeker combinations.

I address both challenges using an adapted over-sampling strategy.[3] I randomly select a ten-day time window and consider all available vacancies open during that window. I combine these vacancies with all recommended jobseekers and randomly sample from the non-recommended jobseekers registered at the same time to achieve the desired outcome mean. I discard vacancy-job seeker pairs where the job seeker's regional work preference does not align with the vacancy's place of work. I opt for a conservative outcome mean

[3]Over-sampling is a common strategy for dealing with imbalances in the outcome variable (Menardi and Torelli, 2014; Japkowicz and Stephen, 2002; Estabrooks, Jo, and Japkowicz, 2004).

of 0.1. I repeat this sampling procedure 20 times. To evaluate the performance of the recommender on hold-out data, I use this approach to construct two separate data sets by sampling from January 2021 to December 2022 for the training sample and March to September 2023 for the hold-out sample.[4]

2.4 Descriptive statistics of matched sample

			Recommended Outcome mean=0.101			*Hiring* Outcome mean=0.091		
	Avg.	St.Dev.	No	Yes	Diff.	No	Yes	Diff.
Age	41.308	12.288	41.431	40.206	-1.225*	41.321	39.850	-1.471*
Occupational (minimum) distance	3.030	2.073	3.279	0.814	-2.465*	3.050	0.956	-2.093*
Age requirement	0.965	0.184	0.963	0.981	0.018*	0.965	0.982	0.017*
Sufficient education	0.896	0.306	0.892	0.925	0.033*	0.895	0.934	0.038*
Sunday match	0.095	0.293	0.084	0.193	0.109*	0.094	0.151	0.057*
Shift work match	0.076	0.265	0.068	0.143	0.075*	0.076	0.129	0.053*
Night work match	0.010	0.097	0.008	0.025	0.017*	0.009	0.023	0.013*
Work from home match	0.000	0.011	0.000	0.000	0.000*	0.000	0.000	0.000*
Workload match	0.580	0.494	0.574	0.629	0.055*	0.579	0.647	0.068*
Car license	0.953	0.211	0.950	0.983	0.033*	0.953	0.973	0.020*
Motor vehicle license	0.999	0.031	0.999	0.999	0.000*	0.999	0.999	0.000
Heavy vehicle license	0.997	0.056	0.997	0.999	0.002*	0.997	0.997	0.000
Sufficient German	0.842	0.365	0.835	0.903	0.068*	0.841	0.886	0.044*
Sufficient English	0.948	0.221	0.945	0.976	0.030*	0.948	0.977	0.029*
Sufficient French	0.933	0.251	0.930	0.958	0.028*	0.932	0.965	0.032*
Sufficient Italian	0.969	0.172	0.967	0.989	0.021*	0.969	0.990	0.021*

Notes: The table reports summary statistics for selected variables in the matched vacancy-job seeker sample. The sample size is $N = 580\,885$. The columns labeled 'Avg.' and 'St.Dev' report the average and standard deviation. The next three columns report the averages for recommended and non-recommended vacancy-jobseeker pairs, as well as their differences. Finally, the last three columns report averages for pairs resulting in a job match, pairs without a job match, and their difference in means.

Table 1: Vacancy-jobseeker characteristics in the matched sample

Table 1 provides summary statistics for selected variables in the generated vacancy-jobseeker sample. The selected variables indicate whether there is a match between vacancy and jobseeker in terms of age, education, language, and work type requirements. We furthermore add the occupational distance measure proposed by Klaeui et al. (2023) measuring the skills overlap between occupations. In addition to sample means and standard deviations by variable, the table reports differences in means for recommended and non-recommended pairs and for hirings. The results indicate that the match indicators

[4] I exclude January and February 2023 to reduce the risk of data leakage between training and hold-out samples that may arise from jobseekers appearing in both data partitions.

positively correlate with observed caseworker recommendations.

3 Recommdations and hirings

In this section, I explore whether caseworkers' recommendations and hiring are determined by sensitive characteristics. I focus on gender and nationality (Swiss vs. non-Swiss) and also consider the length of job search, which serves as a proxy for long-term unemployment. With respect to the latter, I distinguish between job seekers searching for less than a year and job seekers who are in the second year of their job search.

I estimate the partially linear regression models

$$A_{ij} = \alpha S_j + f(X_{ij}) + \varepsilon_{ij} \qquad \text{for all } (i,j) \in \mathcal{I}, \tag{1}$$

$$\text{and} \qquad Y_{ij} = \beta S_j + f(X_{ij}) + \varepsilon_{ij} \qquad \text{for all } (i,j) \in \mathcal{I} \text{ and } A_{ij} = 1, \tag{2}$$

where the set $\mathcal{I}$ includes all vacancy-jobseeker pairs that are registered at the same time and where jobseekers' regional work preferences align with the vacancy's region of work. The variable A_{ij} is an indicator equal to 1 if a caseworker recommends candidate j for vacancy i, 0 otherwise. Similarly, Y_{ij} is 1 if the employer associated with vacancy i reports back that the vacancy has been filled with candidate j, 0 otherwise. S_j is jobseeker j's protected group, which is set to one if j is either female, a long-term job seeker, or has non-Swiss nationality. X_{ij} collects the full set of vacancy, jobseeker, and pair-specific characteristics. Due to the large number of characteristics and since interactions are likely to play a role, I estimate the models by Double Machine Learning (DML; Chernozhukov et al., 2018), which flexibly absorbs the influence of the control variables using supervised machine learners. I pair DML with the short stacking approach of Ahrens et al. (2024) that allows combining several candidate machine learners through model averaging. The considered learners are the cross-validated lasso, ridge, and three types of gradient-boosted trees as candidate learners (see Table 2 for details).

Results are shown in Table 2. The outcome variable in Columns (1)-(3) is caseworker recommendations. Columns (4)-(6) use hirings as the outcome and restrict the sample to

	Recommendations			*Job matches*		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Gender						
Female	-0.0148^{***}	-0.0147^{***}	-0.0062^{***}	-0.0125^{***}	-0.0087^{**}	-0.0028
	(0.0009)	(0.0007)	(0.0009)	(0.0026)	(0.0028)	(0.0038)
Obs.	494816	494816	494816	49826	49826	49826
Panel B. Nationality						
Non-Swiss	0.0403^{***}	0.0100^{***}	0.0045^{***}	-0.0018	-0.0108^{***}	-0.0077^{*}
	(0.0008)	(0.0008)	(0.0009)	(0.0028)	(0.0030)	(0.0039)
Obs.	494816	494816	494816	49826	49826	49826
Panel C. Length of job search spell						
Longterm	-0.0183^{***}	-0.0139^{***}	-0.0084^{***}	-0.0176^{***}	-0.0153^{***}	-0.0066
	(0.0010)	(0.0009)	(0.0008)	(0.0032)	(0.0032)	(0.0038)
Obs.	448193	448193	448193	46559	46559	46559
Sample		All pairs			Only recommended	
Controls	No	Yes	Yes	No	Yes	Yes
Estimator	OLS	OLS	DDML	OLS	OLS	DDML

Notes: The table shows results from regressing caseworker recommendations (Columns 1-3) or hirings (Columns 4-6) against indicators for female, long-term job seeker, and Swiss nationality. Columns (1) and (4) OLS without controls. Columns (2) and (5) use OLS controlling for caseworker, jobseeker, and combined characteristics. Columns (3) and (6) employ DDML with short-stacking (Chernozhukov et al., 2018; Ahrens et al., 2024). The considered candidate learners are cross-validated lasso, cross-validated ridge, and gradient-boosted trees with low (up to 1,000 trees, early stopping after 10 rounds, maximum tree depth of 5), medium (maximum tree depth of 10), and high regularization (maximum tree depth of 20). I use 5 cross-fitting folds.

Table 2: Unconditional and conditional gaps in recommendations and hirings

all vacancy-jobseeker pairs where the caseworker expressed a recommendation. The OLS estimation without controls in Panel A, Column (1) indicates that the recommendation share for female job seekers is 1.48 percentage points (ppts) lower than for male job seekers. The point estimate changes little when controlling for vacancy and jobseeker characteristics linearly but drops in absolute size to –0.62 ppts. when flexibly adjusting for controls using DDML. Panel B and C indicate that, conditional on their characteristics, non-Swiss are 0.45 more likely to be recommended, while long-term job seekers exhibit an unexplained disadvantage in recommendation shares of 0.8 ppts. The DDML results in Column (6) provide no statistical evidence for the existence of a hiring penalty for women and long-term job seekers. There is a statistically significant penalty for non-Swiss job seekers amounting to 0.8 ppts.

4 Methodology

4.1 Setting

The recommender system constitutes a decision rule $\pi(W_{ij}) \in \Pi$ that determines whether a candidate j's profile is forwarded to vacancy i's owner based on observable jobseeker and vacancy characteristics W_{ij}. The ideal policy rule maximizes the number of hirings occurring through the system's recommendations, i.e.,

$$Q(\pi) = \sum_{(i,j)\in\mathcal{I}} E[Y_{ij}(\pi(W_{ij}))].$$

$Y_{ij}(0)$ and $Y_{ij}(1)$ are the potential hiring outcomes if candidate j is recommended for vacancy i, respectively. I take the perspective of the RAVs and the aim is to maximize their efficacy in creating job matches through recommendations. I thus set $Y_{ij}(0) = 0$, implying that a job match without a recommendation has no value to the decision maker.[5]

4.2 Optimization problem

To estimate $\pi(W_{ij})$, I leverage historical data $\{Y_{ij}, A_{ij}, W_{ij}\}_{(i,j)\in\mathcal{I}}$, which include observed job matches Y_{ij}, vacancy and jobseeker characteristics W_{ij}, and past caseworker recommendations, denoted by $A_{ij} \in \{0, 1\}$. Since $Y_{ij}(1)$ is unobserved, it is not possible to evaluate $Q(\pi)$ directly, but I estimate the value of candidate policy π using augmented inverse-propensity weighted scores (Robins, Rotnitzky, and Zhao, 1994; Athey and Wager, 2021):

$$\widehat{Q}(\pi(W_{ij})) = \frac{1}{|\mathcal{I}|} \sum_{(i,j)\in\mathcal{I}} \left(\frac{Y_i - \hat{\mu}^{-k}(W_{ij})}{\hat{e}^{-k}(W_{ij})} \mathbb{1}\{\pi(W_{ij}) = 1\} + \hat{\mu}^{-k(i)}(W_{ij}) \right). \tag{3}$$

Here, $\hat{\mu}^{-k}$ and $\hat{e}^{-k}$ are cross-fitted estimates of $\mu(W_{ij}) = E[Y_{ij}|W_{ij}, A_{ij} = 1]$ and $e(W_{ij}) = E[A_{ij}|W_{ij}]$, respectively. The "$-k$" superscript indicates that the predicted values are

[5]From the perspective of the job seeker, it likely makes no difference whether the hiring is due to a recommendation or not. However, here, the aim is to construct a recommender system that maximizes its success in creating job matches. A practical reason for focusing on RAV efficacy rather than overall hirings is that hirings are only observed if a recommendation has been filed.

obtained by randomly splitting the data in K folds and fitting a nonparametric estimator on all folds but fold k, which is the fold that vacancy-jobseeker pair (i, j) falls into. The use of cross-fitting avoids estimation errors from estimating the conditional expectation functions carrying over to the estimation of the policy rule (Athey and Wager, 2021).

This allows us to define the constrained maximization problem as

$$\hat{\pi} = \arg\max_{\pi \in \Pi} \widehat{Q}(\pi(W_{ij})) \tag{4}$$

$$\text{s.t.} \qquad \frac{1}{|\mathcal{I}|} \sum_{(i,j) \in \mathcal{I}} \pi(W_{ij}) \leq \delta \tag{5}$$

$$F(\pi(W_{ij})) \leq \phi \tag{6}$$

The constraint (5) defines an upper bound for the share of vacancy-jobseeker pairs that are associated with a recommendation. In the application, I set δ to 0.1, the observed recommendation rate in the sampled data. The rationale for defining an upper bound is that recommendations are costly to process for employers, and recommending everyone would otherwise be a trivial solution. The constraint (6) accommodates fairness constraints and is discussed below.

With respect to the policy class Π, I consider rules of the form

$$\pi(W_{ij}) = \mathbb{1}\{\mu(W_{ij}) \geq (1 - S_j)\tau_0 + S_i \tau_1\}, \tag{7}$$

where τ_0 and τ_1 are thresholds that are specific to the group membership. The rule assigns the treatment based on the expected job match probability if a recommendation is expressed but allows for differential treatment by group membership. Since $\mu(W_{ij})$ is unobserved, it is replaced with out-of-sample predicted values.

4.3 Imposing fairness constraints

The observable vacancy and jobseeker characteristics, $W_{ij} = (X_{ij}, S_j)$, include information relevant to the match quality (e.g., occupation, experience, qualifications), which are collected in the vector X_{ij}, but also the jobseeker's protected, typically immutable group membership $S_j \in \{0,1\}$. Legal and ethical considerations may require to treat these protected characteristics differently. The fairness literature has suggested several notions of fairness (Castelnovo et al., 2022; Barocas, Hardt, and Narayanan, 2023). I focus here on two types of fairness.

Demographic parity (DR) requires that the recommendation $\pi(W_{ij})$ are independent of sensitive attribute S_j. Formally,

$$P(\pi(W_{ij}) = 1|S_j = 0) = P(\pi(W_{ij}) = 1|S_j = 1).$$

DR can be enforced by setting

$$F(\pi(W_{ij})) = \frac{\sum_{(i,j)\in\mathcal{I}} \pi(W_{ij})S_j}{\sum_{(i,j)\in\mathcal{I}} S_j} - \frac{\sum_{(i,j)\in\mathcal{I}} \pi(W_{ij})(1-S_j)}{\sum_{(i,j)\in\mathcal{I}}(1-S_j)}.$$

A disadvantage of demographic parity is that it is blind towards the probability of a job match.

Predictive parity (PP) requires that the expected hiring given a recommendation is the same across groups, i.e.,

$$P(Y_{ij}|\pi(W_{ij}) = 1, S_j = 0) = P(Y_{ij}|\pi(W_{ij}) = 1, S_j = 1).$$

Predictive parity can be enforced by setting

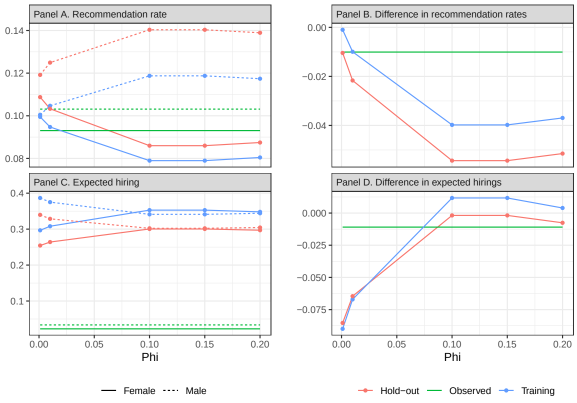
$$F(\pi(W_{ij})) = \frac{\sum_{(i,j)\in\mathcal{I}} \mu(W_{ij})S_j\pi(W_{ij})}{\sum_{(i,j)\in\mathcal{I}} S_j\pi(W_{ij})} - \frac{\sum_{(i,j)\in\mathcal{I}} \mu(W_{ij})(1-S_j)\pi(W_{ij})}{\sum_{(i,j)\in\mathcal{I}}(1-S_j)\pi(W_{ij})}.$$

5 Results

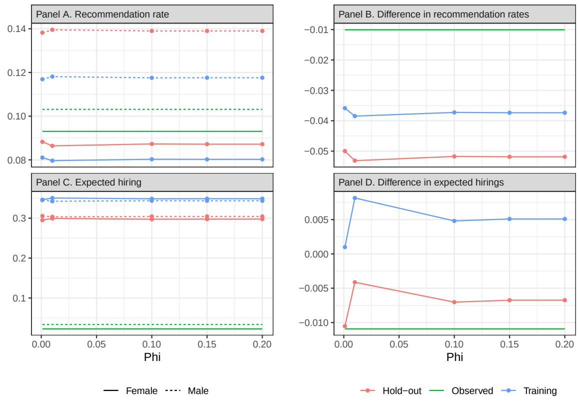
5.1 Fairness evaluation

In Figure 1, I report the recommendation rate and expected hiring rate when enforcing demographic parity (left-hand side) or predictive parity (right-hand side) across gender, nationality, and job search duration, respectively (see also Table 4). For example, Figure 1a reports results from a recommender system enforcing demographic parity across gender. Panel A shows the recommendation rate for male and female job seekers as a function of ϕ, which determines how strictly fairness constraints are enforced. Panel B visualizes the differences in recommendation shares between genders. The second row in Figure 1a reports the expected hiring rate by gender (Panel C) and the gender gap in expected hiring (Panel D). Figure 1b follows the same structure but is based on a recommender system enforcing predictive parity across gender. Figures 1c–1d and Figures 1e–1f show results when enforcing fairness constraints by Swiss vs. non-Swiss nationality and by short vs. long-term jobseekers, respectively. The results are separately reported for training and hold-out samples. Finally, for reference, all graphs also show the caseworkers' recommendation rates and expected hiring in the hold-out sample.

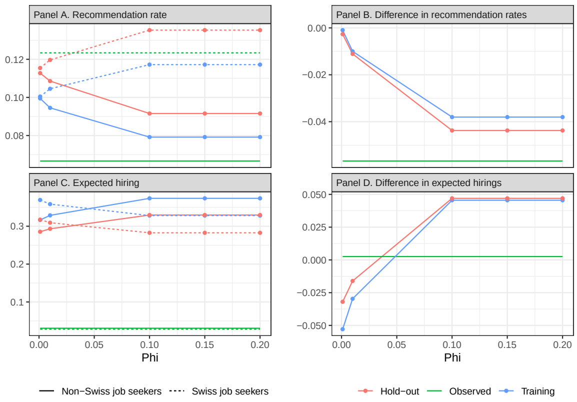
Gender. Focusing first on the gender dimension, Panel A in Figure 1a shows that a recommendation system without binding fairness constraints yields a gap in recommendation shares of 3.8 ppt. in the training and 5.2 ppt. in the hold-out sample. Specifically, the fairness-unconstrained recommender system expresses candidate recommendations of male job seekers with a probability of 13.9% in the hold-out sample, while only 8.8% of female job seekers are recommended. For comparison, the observed recommendation rates of male and female job seekers are 10.3% and 9.3%, suggesting that a recommender system may aggravate biases against a particular group compared to observed outcomes. However, the gap in recommendation rates can be reduced by enforcing demographic parity. In Figure 1a, the gap in recommendation rates is reduced to approximately zero in the training and 1.0 ppt. in the hold-out sample when demographic parity is strictly enforced (i.e., $\phi = 0.001$). At the same time, enforcing demographic parity increases the gender gap in expected hiring rates from 0.8 to 8.5 ppt in the hold-out data (see Panels C and D). Figure 1b shows that enforcing predictive parity with respect to gender has little impact on recommendation shares and expected hiring rates, suggesting that predictive parity is approximately satisfied without imposing constraints.



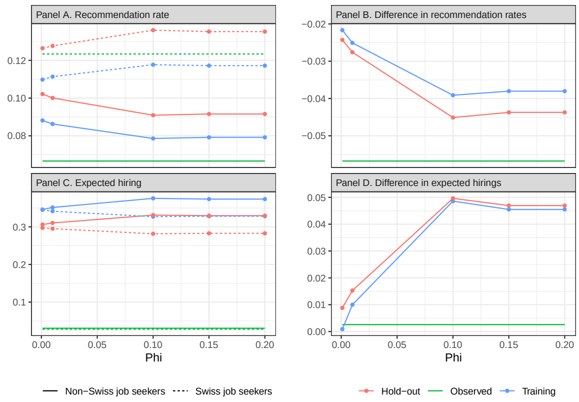
(a) Demographic parity across gender



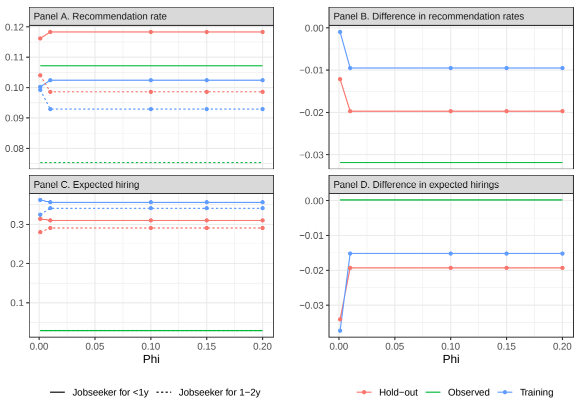
(b) Predictive parity across gender



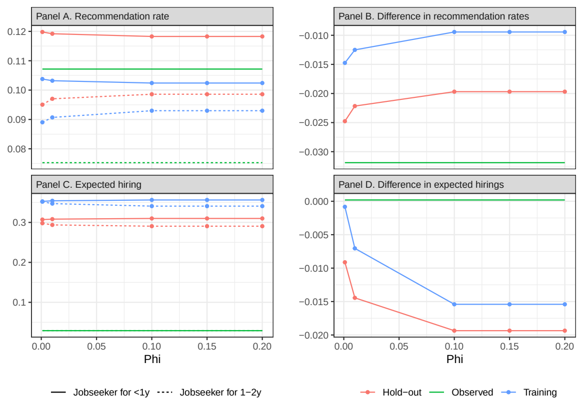
(c) Demographic parity across nationality



(d) Predictive parity across nationality



(e) Demographic parity across search duration



(f) Predictive parity across search duration

Notes: The figures report the recommendation rate (top-left panel), difference in recommendation rates (top-right panel), expected hiring rate (bottom-left panel), and gap in expected hiring rate (bottom right panel) as a function of ϕ, which determines how strictly fairness constraints are enforced. Figure 1a and 1b enforce demographic parity and predictive parity by gender, respectively. Figure 1c–1d and Figure 1e–1f focus on Swiss vs. non-Swiss nationality, and short vs. long-term job search instead. Results are reported separately for the training and hold-out sample, as well as for the observed caseworker recommendations.

Figure 1: Recommendation share and expected hirings for different fairness-constraint recommendation systems

Nationality. Next, I turn to the recommender systems' treatment of Swiss and non-Swiss job seekers. In the hold-out sample, caseworkers are 5.7 ppt. less likely to recommend non-Swiss compared to Swiss job seekers. This gap is only slightly smaller in absolute size when considering recommendations from a fairness-unconstrained recommender system (4.4 ppt). Imposing demographic parity reduces the recommendation penalty for non-Swiss job seekers to close to zero. While there is no evidence that caseworkers produce a gap in expected hiring between Swiss and non-Swiss, the fairness-unconstrained recommender system yields an advantage for non-Swiss over Swiss job seekers, amounting to 4.7 ppt. This advantage flips to a penalty of 3.2 ppt when enforcing demographic parity and reduces to close to zero when imposing predictive parity.

Duration of job search. Finally, I consider differences in recommendations between job seekers who have been on a job search for up to 1 year (here referred to as 'short-term job seekers') and job seekers who are in the second year of their job search ('long-term job seekers'). Long-term job seekers are 3.2 ppt. less likely to be recommended by caseworkers in the hold-out sample, while there is no gap in expected hirings. The recommender system without fairness constraints, however, yields gaps in recommendation rates and expected hirings of around 1.9 ppt. in favor of short-term job seekers. Enforcing demographic parity reduces the gap in recommendation rates to approximately zero in the training and 1.2 ppt. in the hold-out sample while increasing the absolute gap in expected hirings to 3.4 ppt. On the other side, imposing predictive parity reduces the gap in expected hiring to 0.9 ppt. while increasing the gap in recommendation rates to 2.5 ppt.

Summary. To summarize, there is evidence that caseworkers' recommendations approximately comply with a predictive parity notion of fairness. This is supported by the discrimination test in Table 2, where the magnitude of the unexplained gaps in recommendations is economically small, albeit statistically significant. However, an unconstrained recommender system may fail to comply with both demographic parity and predictive parity. A constrained recommender system can ensure compliance with either demographic parity or predictive parity, but not both. There is empirical evidence for a trade-off in that enforcing demographic parity may increase gaps in expected hiring, and enforcing predictive parity tends to lead to a rise in the gap in recommendation rates. Finally, enforcing either demographic or predictive parity does not affect the overall expected hiring rate compared to a fairness-unconstrained recommender system. Compared to the caseworker recommendations, all recommender systems generate substantially higher expected hirings per recommended job seeker.

5.2 Adding concentration constraints

Next, I examine the distribution of recommendations across vacancies. A concern is that the recommendation system concentrates candidate recommendations among a few vacancies that are attractive to a large pool of candidates. Since vacancies are filled by only one candidate, a high degree of concentration would undermine the efficacy of the recommender system in creating job matches.

Table 3 shows the distribution of the number of recommendations across vacancies for caseworkers' recommendations and different recommender systems. Only around 20.0% of vacancies receive no recommendations by caseworkers, and 90.0% are associated with 5 or fewer recommendations. In contrast, the fairness-unconstrained recommender system does not assign any recommendations to between 27.7 and 29.6% of vacancies, suggesting that fewer firms benefit from candidate recommendations. The Gini coefficients, reported in the same table, confirm that the recommender systems' recommendations are more concentrated than caseworker recommendations. The recommender systems' Gini coefficients are around 0.63 in the hold-out sample, compared to around 0.51 for caseworker recommendations. The distribution of recommendations is approximately unchanged when demographic parity or predictive parity is enforced (see columns labeled 'Fairness constraint').

To control the issue of concentration, I consider an extended decision rule where the effective threshold depends not only on group membership but also on the rank of the

	Observed	Recommender system				
		Unconstr.	Fairness constraint		Concentration constraint	
			DP	PP	DP	PP
Panel A. Gender						
No recommendations	0.2005	0.2803	0.2744	0.2808	0.1429	0.1694
Less than 1	0.4191	0.5083	0.5074	0.5065	0.3601	0.3659
Less than 2	0.6809	0.6602	0.6553	0.6598	0.5552	0.5381
Less than 3	0.7643	0.7598	0.7594	0.7616	0.7084	0.6809
Less than 4	0.8644	0.8229	0.8297	0.8229	0.7999	0.7895
Less than 5	0.8896	0.8680	0.8693	0.8684	0.8693	0.8508
Less than 10	0.9860	0.9527	0.9554	0.9518	0.9743	0.9748
Less than 15	0.9968	0.9811	0.9815	0.9811	0.9932	0.9982
Gini	0.5128	0.6386	0.6330	0.6383	0.4878	0.4924
Panel B. Nationality						
No recommendations	0.2005	0.2776	0.2821	0.2794	0.1442	0.1537
Less than 1	0.4191	0.5074	0.5088	0.5065	0.3556	0.3610
Less than 2	0.6809	0.6616	0.6620	0.6575	0.5561	0.5444
Less than 3	0.7643	0.7544	0.7621	0.7575	0.7134	0.6954
Less than 4	0.8644	0.8224	0.8242	0.8224	0.8080	0.7895
Less than 5	0.8896	0.8675	0.8671	0.8666	0.8648	0.8598
Less than 10	0.9860	0.9527	0.9554	0.9509	0.9730	0.9743
Less than 15	0.9968	0.9806	0.9806	0.9811	0.9932	0.9964
Gini	0.5128	0.6389	0.6381	0.6381	0.4877	0.4882
Panel C. Length of job search spell						
No recommendations	0.2183	0.2959	0.2986	0.2936	0.1673	0.1195
Less than 1	0.4371	0.5309	0.5327	0.5313	0.3839	0.3500
Less than 2	0.6955	0.6861	0.6874	0.6843	0.5819	0.5765
Less than 3	0.7749	0.7781	0.7803	0.7763	0.7280	0.7388
Less than 4	0.8724	0.8408	0.8412	0.8408	0.8227	0.8354
Less than 5	0.8981	0.8764	0.8791	0.8755	0.8796	0.8868
Less than 10	0.9874	0.9603	0.9599	0.9590	0.9824	0.9824
Less than 15	0.9968	0.9829	0.9838	0.9833	0.9986	0.9982
Gini	0.5216	0.6431	0.6441	0.6429	0.4880	0.4560

Notes: The table reports the share of vacancies with no or fewer than 1–15 recommendations in the hold-out sample. The column 'Observed' refers to caseworkers' recommendations. The column 'Unconstr.' reports shares for a fairness-unconstrained recommender system. The next two columns impose demographic parity (DP) or predictive parity (PP) with $\phi = 0.001$. The last two columns add the congestion constraint in (9). The table also reports the Gini coefficient for the respective recommender systems.

Table 3: Number of recommendations per vacancy

candidate. Specifically,

$$\pi(W_{ij}) = \mathbb{1}\{\mu(W_{ij}) \geq (1 - S_j)\tau_0 + S_i\tau_1 + R(\mu(W_{ij}))\tau_2 + R(\mu(W_{ij}))\tau_3\}. \quad (8)$$

Here, $R_0(\mu(W_{ij}))$ denotes the rank of candidate j for vacancy i.[6] For example, a positive value of τ_2 implies that the effective threshold is higher for each consecutive recommendation. If $\tau_3 < 0$, then the effective threshold increases at a decreasing rate. Furthermore, I impose the additional constraint that the Gini coefficient over the number of recommendations per vacancy is smaller than a to-be-defined threshold, i.e.,

$$G(\pi(W_{ij})) \leq \delta. \quad (9)$$

I set δ to the observed Gini coefficient in the training sample.

The last two columns of Table 3 show results of recommender systems that impose both fairness constraints and constraints on the concentration of candidate recommendation. The results show that this approach successfully addresses the congestion concern. The Gini coefficients decreases to around 0.49. The share of vacancies without candidate recommendations drops from 27.4–29.9% to between 14.4% and 16.9%, depending on the fairness constraint and fairness dimension.

Table 4 returns to the question of fairness. The table compares recommendation rates and expected hiring of recommender systems with and without fairness constraints and with and without concentration constraints. The columns labeled ‘Fairness constraint’ show results when enforcing demographic parity or predictive parity. The columns ‘Concentration constraint’ enforce, in addition, an upper bound on the Gini concentration as described above. The results suggest that this additional constraint has no implications for the relative fairness metrics. However, overall expected hiring drops by around 2 ppts, suggesting that requiring less concentrated recommendations leads to lower average hiring chances per recommended candidate.

[6] To facilitate interpretability, I standardized the rank such that $R_0(\mu(W_{ij})) = 0$ corresponds to the first-ranked candidate. Hence, τ_0 and τ_1 are the effective thresholds for the candidate ranked first.

	Observed	Recommender system				
		Unconstr.	Fairness constraint		Concentration constraint	
			DP	PP	DP	PP
Panel A. Gender						
Share recommended	0.0985	0.1152	0.1144	0.1152	0.1202	0.1205
Female	0.0930	0.0875	0.1087	0.0883	0.1161	0.0926
Male	0.1031	0.1389	0.1192	0.1382	0.1238	0.1445
Difference	−0.0101	−0.0515	−0.0104	−0.0500	−0.0078	−0.0518
Expected hiring	0.0288	0.3019	0.3022	0.3019	0.2835	0.2804
Female	0.0226	0.2969	0.2544	0.2951	0.2346	0.2760
Male	0.0335	0.3045	0.3397	0.3056	0.3228	0.2829
Difference	−0.0109	−0.0076	−0.0854	−0.0105	−0.0882	−0.0069
Panel B. Nationality						
Share recommended	0.0985	0.1161	0.1142	0.1158	0.1200	0.1208
Non-Swiss job seekers	0.0666	0.0916	0.1127	0.1021	0.1181	0.1046
Swiss seekers	0.1234	0.1353	0.1154	0.1264	0.1216	0.1335
Difference	−0.0568	−0.0438	−0.0027	−0.0243	−0.0035	−0.0289
Expected hiring	0.0288	0.2989	0.3036	0.3006	0.2844	0.2809
Non-Swiss job seekers	0.0306	0.3296	0.2855	0.3060	0.2668	0.2895
Swiss seekers	0.0280	0.2827	0.3174	0.2972	0.2978	0.2756
Difference	0.0026	0.0470	−0.0320	0.0088	−0.0309	0.0139
Panel C. Length of job search spell						
Share recommended	0.1015	0.1148	0.1140	0.1154	0.1185	0.1210
Jobseeker for 1-2y	0.0753	0.0986	0.1040	0.0950	0.1181	0.1014
Jobseeker for <1y	0.1072	0.1183	0.1162	0.1198	0.1186	0.1252
Difference	−0.0319	−0.0197	−0.0121	−0.0248	−0.0005	−0.0239
Expected hiring	0.0291	0.3069	0.3084	0.3057	0.2877	0.2833
Jobseeker for 1-2y	0.0293	0.2906	0.2798	0.2980	0.2425	0.2723
Jobseeker for <1y	0.0291	0.3099	0.3139	0.3071	0.2975	0.2853
Difference	0.0002	−0.0193	−0.0341	−0.0091	−0.0550	−0.0130

Notes: The table reports performance and fairness metrics of observed caseworker recommendations, and for five different recommender systems: 'Unconstr.' refers to a recommender system without fairness constraints and without concentration constraints. The two columns labeled 'Fairness constraint' show results when enforcing demographic parity (DP) or predictive parity (PP). The columns 'Concentration constraint' enforce, in addition, an upper bound on the Gini concentration. The three panels focus on three fairness-relevant dimensions: the gender of job seekers (Panel A), Swiss vs. non-Swiss job seekers (Panel B), and short vs. long-term job seekers (Panel C). In each panel, I report the share of recommended overall, by group and the difference between groups. Furthermore, in each panel, I report expected hiring changes by group and overall.

Table 4: Recommendation share and expected hirings for different recommender systems subject to fairness and concentration constraints

6 Conclusion

In this study, I design a recommender system for suggesting job candidates to firms. The recommender system builds on a policy learning framework and is geared towards optimizing expected hirings but can accommodate fairness and concentration constraints.

The focus on fairness is motivated by a large literature highlighting the risk that recommender systems may inadvertently discriminate against individuals with protected group memberships. Indeed, I find that, while historic caseworkers' decisions are approximately in line with a predictive parity notion of fairness, an unconstrained recommender system may violate both demographic and predictive parity. A constrained recommender system can correct the recommender system's failure to comply with either demographic or predictive fairness but cannot enforce compliance with both fairness notions. The results are consistent with an empirical trade-off between the two fairness notions in that enforcing demographic parity requires, for example, the recommender system to recommend female job seekers with a lower expected hiring chance. However, imposing either notion of fairness does not lead to a drop in expected hiring. Results from this study highlight that notions of fairness should be explicitly considered and enforced when designing a recommender system.

The results also reveal that recommender systems tend to increase the concentration of recommendations across vacancies, likely undermining the efficacy of the system in creating jobs. The policy learning framework is, however, sufficiently flexible in that it can accommodate additional Gini-based constraints, which successfully control the congestion of recommendations. While enforcing an upper bound on concentration does reduce expected hiring, the recommender system does outperform caseworkers' performance in terms of expected hiring chances. Imposing concentration constraints also does not conflict with the recommender systems' ability to control fairness.

References

Ahrens, Achim, Patrick Arni, Dominik Hangartner, Rafael Lalive, Tobias Lehmann, and Joëlle Pianzola (2021). *Monitoringevaluation der Stellenmeldepflicht I.* Tech. rep. Grundlagen für die Wirtschaftspolitik Nr. 21. Staatssekretariat für Wirtschaft SECO, Bern.

Ahrens, Achim, Christian B. Hansen, Mark E. Schaffer, and Thomas Wiemann (2024). *Model Averaging and Double Machine Learning.* arXiv:2401.01645 [econ, stat]. URL: http://arxiv.org/abs/2401.01645.

Angwin, Julia, Jeff Larson, Surya Mattu, and Lauren Kirchner (2016). *Machine bias.* Tech. rep. ProPublica.

Arnold, David, Will S. Dobbie, and Peter Hull (2024). *Building Non-Discriminatory Algorithms in Selected Data.* Working Paper. URL: https://www.nber.org/papers/w32403.

Athey, Susan and Stefan Wager (2021). "Policy Learning With Observational Data". *Econometrica* 89.1, pp. 133–161.

Bamert, Justus, Andreas Beerli, Boris Kaiser, Daniel Kopp, and Michael Siegenthaler (2021). *Monitoringevaluation der Stellenmeldepflicht I.* Tech. rep. Grundlagen für die Wirtschaftspolitik Nr. 19. Staatssekretariat für Wirtschaft SECO, Bern.

Bansak, Kirk, Jeremy Ferwerda, Jens Hainmueller, Andrea Dillon, Dominik Hangartner, Duncan Lawrence, and Jeremy Weinstein (2018). "Improving refugee integration through data-driven algorithmic assignment". *Science* 359.6373, pp. 325–329.

Barbanchon, Thomas Le, Lena Hensvik, and Roland Rathelot (2023). *How can AI improve search and matching? Evidence from 59 million personalized job recommendations.* Preliminary version. IAB Special Lecture with Roland Rathelot (CREST).

Barocas, Solon, Moritz Hardt, and Arvind Narayanan (2023). *Fairness and Machine Learning: Limitations and Opportunities.* MIT Press.

Behaghel, Luc, Sofia Dromundo, Marc Gurgand, Yagan Hazard, and Thomas Zuber (2024). "The Potential of Recommender Systems for Directing Job Search: A Large-Scale Experiment". *SSRN Electronic Journal.*

Bied, Guillaume, Christophe Gaillac, Morgane Hoffmann, Philippe Caillou, Bruno Crépon, Solal Nathan, and Michele Sebag (2023). "Fairness in job recommendations: estimating, explaining, and reducing gender gaps". In: *AEQUITAS 2023-First AEQUITAS Workshop on Fairness and Bias in AI| co-located with ECAI 2023*. Vol. 3523. CEUR-WS. org.

Bied, Guillaume, Elia Perennes, Victor Alfonso Naya, Philippe Caillou, Bruno Crépon, Christophe Gaillac, and Michele Sebag (2022). "Congestion-Avoiding Job Recommendation with Optimal Transport". In: URL: https://inria.hal.science/hal-03540316.

Braun-Dubler, Nils, Tabea Kaderli, Vera Frei, Leonardo Longhini, and Patrick Puhani (2021). *Monitoringevaluation der Stellenmeldepflicht I.* Tech. rep. Grundlagen für die Wirtschaftspolitik Nr. 20a. Staatssekretariat für Wirtschaft SECO, Bern. URL: https://www.seco.admin.ch/dam/seco/de/dokumente/Publikationen_Dienstleistungen/Publikationen_Formulare/Arbeit/Arbeitsmarkt/Arbeitsmarktforschung/monitoringevaluation_stellenmeldepflicht_2_wp20.pdf.download.pdf/Schlussbericht_Monitoringevaluation_IWSB_FINAL_20.pdf.

Castelnovo, Alessandro, Riccardo Crupi, Greta Greco, Daniele Regoli, Ilaria Giuseppina Penco, and Andrea Claudio Cosentini (2022). "A clarification of the nuances in the fairness metrics landscape". *Scientific Reports* 12.1. Publisher: Nature Publishing Group, p. 4209.

Chernozhukov, Victor, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins (2018). "Double/debiased machine learning for treatment and structural parameters". *The Econometrics Journal* 21.1. tex.ids= Chernozhukov2018a publisher: John Wiley & Sons, Ltd (10.1111), pp. C1–C68.

Dressel, Julia and Hany Farid (2018). "The accuracy, fairness, and limits of predicting recidivism". *Science Advances* 4.1, eaao5580.

Dwork, Cynthia, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Rich Zemel (2011). *Fairness Through Awareness*. arXiv:1104.3913 [cs]. URL: http://arxiv.org/abs/1104.3913.

Estabrooks, Andrew, Taeho Jo, and Nathalie Japkowicz (2004). "A Multiple Resampling Method for Learning from Imbalanced Data Sets". *Computational Intelligence* 20.1. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.0824-7935.2004.t01-1-00228.x, pp. 18–36.

Hangartner, Dominik, Daniel Kopp, and Michael Siegenthaler (2021). "Monitoring hiring discrimination through online recruitment platforms". *Nature* 589.7843. Publisher: Nature Publishing Group, pp. 572–576.

Hardt, Moritz, Eric Price, Eric Price, and Nati Srebro (2016). "Equality of Opportunity in Supervised Learning". In: *Advances in Neural Information Processing Systems*. Vol. 29. Curran Associates, Inc. URL: `https://proceedings.neurips.cc/paper/2016/hash/9d2682367c3935defcb1f9e247a97c0d-Abstract.html`.

Japkowicz, Nathalie and Shaju Stephen (2002). "The class imbalance problem: A systematic study". *Intell. Data Anal.* 6.5, pp. 429–449.

Kim, Kwangho and José R. Zubizarreta (2023). *Fair and Robust Estimation of Heterogeneous Treatment Effects for Policy Learning*. arXiv:2306.03625 [cs, stat]. URL: `http://arxiv.org/abs/2306.03625`.

Kircher, Philipp (2022). "Job Search in the 21St Century". *Journal of the European Economic Association* 20.6. Number: 6, pp. 2317–2352.

Klaeui, Jeremias, Daniel Kopp, Rafael Lalive, and Michael Siegenthaler (2023). "Labor demand and supply across occupational boundaries". *University of Lausanne and ETH Zurich*.

Kleinberg, Jon, Jens Ludwig, Sendhil Mullainathan, and Ashesh Rambachan (2018). "Algorithmic Fairness". *AEA Papers and Proceedings* 108, pp. 22–27.

Kusner, Matt J., Joshua R. Loftus, Chris Russell, and Ricardo Silva (2018). *Counterfactual Fairness*. arXiv:1703.06856 [cs, stat]. URL: `http://arxiv.org/abs/1703.06856`.

Li, Danielle, Lindsey R. Raymond, and Peter Bergman (2020). *Hiring as Exploration*. Working Paper. URL: `https://www.nber.org/papers/w27736`.

Lippens, Louis, Siel Vermeiren, and Stijn Baert (2023). "The state of hiring discrimination: A meta-analysis of (almost) all recent correspondence experiments". *European Economic Review* 151, p. 104315.

Ludwig, Jens and Sendhil Mullainathan (2021). "Fragile Algorithms and Fallible Decision-Makers: Lessons from the Justice System". *Journal of Economic Perspectives* 35.4, pp. 71–96.

Manski, Charles F. (2004). "Statistical Treatment Rules for Heterogeneous Populations". *Econometrica* 72.4, pp. 1221–1246.

— (2007). "Minimax-regret treatment choice with missing outcome data". *Journal of Econometrics*. Endogeneity, instruments and identification 139.1, pp. 105–115.

Menardi, Giovanna and Nicola Torelli (2014). "Training and assessing classification rules with imbalanced data". *Data Mining and Knowledge Discovery* 28.1, pp. 92–122.

Mueller, Andreas I and Johannes Spinnewijn (2023). *The Nature of Long-Term Unemployment: Predictability, Heterogeneity and Selection*. Working Paper 30979. National Bureau of Economic Research. URL: `http://www.nber.org/papers/w30979`.

Neumark, David (2018). "Experimental Research on Labor Market Discrimination". *Journal of Economic Literature* 56.3, pp. 799–866.

Robins, James M., Andrea Rotnitzky, and Lue Ping Zhao (1994). "Estimation of Regression Coefficients When Some Regressors are not Always Observed". *Journal of the American Statistical Association* 89.427, pp. 846–866.

Rus, Clara, Jeffrey Luppes, Harrie Oosterhuis, and Gido H. Schoenmacker (2022). *Closing the Gender Wage Gap: Adversarial Fairness in Job Recommendation*. arXiv:2209.09592 [cs]. URL: `http://arxiv.org/abs/2209.09592`.

Savage, L. J. (1951). "The Theory of Statistical Decision". *Journal of the American Statistical Association* 46.253. Publisher: American Statistical Association, Taylor & Francis, Ltd., pp. 55–67.

Sheldon, George and Conny Wunsch (2021). *Monitoringevaluation der Stellenmeldepflicht II*. Tech. rep. Grundlagen für die Wirtschaftspolitik Nr. 22. Staatssekretariat für Wirtschaft SECO, Bern. URL: `https://www.seco.admin.ch/dam/seco/de/dok` umente/Publikationen_Dienstleistungen/Publikationen_Formulare/Arbeit/Arbeitsmarkt/Arbeitsmarktforschung/wirkungsevaluationstellenmeldepflicht_2_wp22.pdf.download.pdf/Schlussbericht_Wirkungsevaluation_UniBasel_FINAL_22.pdf.

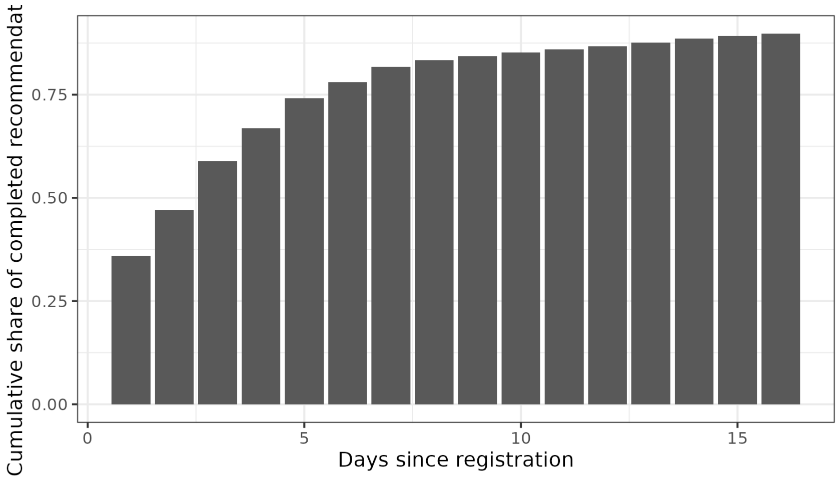
Viviano, Davide and Jelena Bradic (2022). *Fair Policy Targeting.* arXiv:2005.12395 [econ, math, stat]. URL: http://arxiv.org/abs/2005.12395.

Wald, Abraham (1950). *Statistical decision functions.* Wiley.

Zafar, Muhammad Bilal, Isabel Valera, Manuel Gomez Rodriguez, and Krishna P. Gummadi (2017). "Fairness Beyond Disparate Treatment & Disparate Impact: Learning Classification without Disparate Mistreatment". In: *Proceedings of the 26th International Conference on World Wide Web.* arXiv:1610.08452 [cs, stat], pp. 1171–1180. URL: http://arxiv.org/abs/1610.08452.

Online Appendix

A Descriptives



Notes: The table reports the cumulative share of completed recommendations by the number of days since the vacancy's registration date.

Figure A.1: The timing of job recommendations

	Avg.	St.Dev.	Min.	Max.	Not NA
Age preference defined	0.06	0.24	0.00	1.00	290150
German level: very good	0.23	0.42	0.00	1.00	267719
English level: very good	0.03	0.18	0.00	1.00	267719
French level: very good	0.07	0.26	0.00	1.00	267719
Italian level: very good	0.02	0.14	0.00	1.00	267719
Preference for female jobseeker	0.02	0.15	0.00	1.00	290150
Age preference: upper	46.38	9.22	20.00	65.00	16574
Age preference: lower	22.38	5.11	16.00	64.00	13393
Private car requirement	0.10	0.30	0.00	1.00	290150

Notes: The table reports average (column labeled 'Avg.'), standard deviation ('St.Dev.'), minimum ('Min.'), maximum ('Max.') and the number of non-missing observations ('Not NA') for selected variables in the vacancy sample.

Table A.1: Vacancy characteristics

	Share (%)	Count
Employment duration		
Permanent	78.14	226720
Short-term	0.02	44
Temporary	21.85	63386
Total	100.00	290150
Education requirement		
Not specified	54.40	157843
Primary & lower secondary education	13.80	40031
Tertiary education	2.09	6073
Upper secondary education	29.71	86203
Total	100.00	290150
Driving licence requirement		
Car	19.99	58014
None	79.19	229773
Truck	0.81	2363
Total	100.00	290150
Experience requirement		
1 to 3 years	51.06	148156
Less than 1 year	5.28	15320
More than 3 years	29.38	85242
None	2.63	7633
Not specified	11.65	33799
Total	100.00	290150
Qualification requirement		
Not specified	14.06	40792
Semi-skilled	13.53	39255
Skilled	34.91	101288
Unskilled	37.50	108815
Total	100.00	290150

Notes: The table reports shares by category (in %) and counts for selected variables in the vacancy sample.

Table A.2: Vacancy characteristics

	Avg.	St.Dev.	Min.	Max.	Not NA
Age	41.58	12.53	14.00	70.00	6335360
Special work forms	0.95	0.22	0.00	1.00	1508714
Work from home	0.12	0.32	0.00	1.00	1508714
Apprenticeship	0.05	0.21	0.00	1.00	1508714
Night work	0.29	0.46	0.00	1.00	1508714
Shift work	0.69	0.46	0.00	1.00	1508714
Sunday/holiday work	0.62	0.49	0.00	1.00	1508714
Car licence	0.74	0.44	0.00	1.00	6335368
German level: very good	0.42	0.49	0.00	1.00	6335368
English level: very good	0.23	0.42	0.00	1.00	6335368
French level: very good	0.30	0.46	0.00	1.00	6335368
Italian level: very good	0.13	0.34	0.00	1.00	6335368
Female	0.47	0.50	0.00	1.00	6335368
License heavy vehicles	0.05	0.22	0.00	1.00	6335368
Motor license	0.05	0.23	0.00	1.00	6335368

Notes: The table reports average (column labeled 'Avg.'), standard deviation ('St.Dev.'), minimum ('Min.'), maximum ('Max.') and the number of non-missing observations ('Not NA') for selected variables in the job seeker sample.

Table A.3: Jobseeker characteristics

	Share (%)	Count
Experience		
1 to 3 years	13.74	870680
Less than 1 year	2.82	178396
More than 3 years	80.96	5129175
None	2.48	157117
Total	100.00	6335368
Qualification		
Basic	18.07	1144558
None	34.86	2208340
Professional	46.41	2940262
	0.67	42208
Total	100.00	6335368
Residence permit		
B	19.93	1262335
C	26.68	1690490
Citizen	50.54	3201919
Other	2.85	180624
Total	100.00	6335368
Education		
Not specified	4.01	254078
Primary/lower secondary education	23.95	1517313
Upper secondary education	46.42	2940617
Tertiary education	25.62	1623360
Total	100.00	6335368
Mother tongue		
CH-Deutsch	5.36	339867
Deutsch	28.78	1823464
Franzoesisch	21.63	1370224
Italienisch	7.10	449624
Other	37.10	2350461
Raetoromanisch	0.03	1728
Total	100.00	6335368

Notes: The table reports shares by category (in %) and counts for selected variables in the job seeker sample.

Table A.4: Jobseeker characteristics (continued)

B Regression results from linear probability models

	Dependent variable:		
	Recom.	*Job match*	
	(1)	(2)	(3)
Intercept	−0.2183***	−0.0154	0.1078
	(0.0279)	(0.0083)	(0.1165)
Female	−0.0161***	−0.0023***	−0.0077*
	(0.0010)	(0.0004)	(0.0039)
Age above 55	−0.0216***	−0.0026***	−0.0063
	(0.0013)	(0.0005)	(0.0052)
Swiss nationality	−0.0081***	−0.0001	0.0068
	(0.0011)	(0.0004)	(0.0042)
Secondary education	0.0016	−0.0019**	−0.0185***
	(0.0016)	(0.0006)	(0.0055)
Tertiary education	0.0111**	−0.0029*	−0.0335**
	(0.0034)	(0.0012)	(0.0123)
Number of qualifications	0.0004	0.0003	0.0036
	(0.0005)	(0.0002)	(0.0022)
Sufficient education	0.0265***	0.0036***	0.0173*
	(0.0021)	(0.0007)	(0.0082)
Education difference	−0.0023***	0.0000	0.0022
	(0.0003)	(0.0001)	(0.0011)
Qualification for occupation 1	0.1989***	0.0127***	0.0136
	(0.0084)	(0.0032)	(0.0189)
Qualification for occupation 2	0.2666***	0.0108**	−0.0083
	(0.0090)	(0.0033)	(0.0197)
Qualification for occupation 3	0.3616***	0.0211***	0.0061
	(0.0094)	(0.0036)	(0.0204)
Qualification difference	−0.0111***	−0.0011***	0.0008
	(0.0007)	(0.0002)	(0.0032)
Sufficient relevant qualification	0.0815***	0.0080***	−0.0065
	(0.0053)	(0.0019)	(0.0080)
Jobseekers' total experience	−0.0003**	−0.0000	−0.0003
	(0.0001)	(0.0000)	(0.0005)
Jobseeker count experience	0.0019*	0.0004	0.0004
	(0.0008)	(0.0003)	(0.0032)
Experience relevant occupation 1y	−0.2177***	−0.0117***	
	(0.0085)	(0.0032)	
Experience relevant occupation 5y	−0.1594***	−0.0072*	−0.0138
	(0.0087)	(0.0033)	(0.0196)
Experience relevant occupation 10y	−0.0696***	−0.0026	−0.0259
	(0.0063)	(0.0024)	(0.0189)
Experience relevant occupation 20y			−0.0357
			(0.0236)
Experience relevant difference	−0.0015***	0.0001	0.0011
	(0.0004)	(0.0001)	(0.0010)
Sufficient experience	0.0196***	0.0005	−0.0156
	(0.0018)	(0.0007)	(0.0099)
Count search	0.0078***	0.0005*	−0.0040
	(0.0006)	(0.0002)	(0.0026)
Avg occupational distance	−0.0105***	−0.0014***	−0.0012
	(0.0005)	(0.0002)	(0.0014)
Avg occupational distance	0.0025*	0.0023***	0.0240***
	(0.0011)	(0.0004)	(0.0044)
Min occupational distance	0.0023***	0.0006**	0.0036
	(0.0005)	(0.0002)	(0.0025)
Match in sunday/holiday work preference	0.0538***	0.0002	−0.0203***
	(0.0023)	(0.0008)	(0.0050)

Notes: The table reports results from fitting linear probability models on the matched vacancy-jobseeker sample. Model (1) uses caseworkers' recommendations as the dependent variable. Models (2)–(3) use job matches. Model (3) restricts the sample to vacancy-jobseeker pairs that are associated with a caseworker recommendation. Models (2)–(3) reply on the full sample of vacancy-jobseeker pairs.

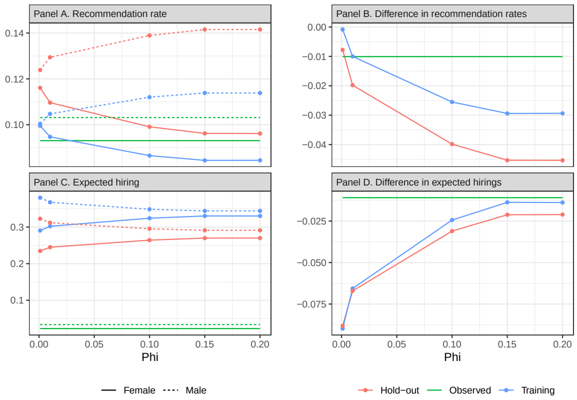
Table B.1: Predicting recommendations and job matches using linear regression

	Dependent variable:		
	Recom.	*Job match*	
	(1)	(2)	(3)
Match in night shift preference	0.0617***	0.0038	−0.0081
	(0.0079)	(0.0029)	(0.0116)
Match in shift work preference	0.0200***	0.0015	−0.0017
	(0.0026)	(0.0010)	(0.0055)
Match in home office preference	0.0733	−0.0101***	−0.0645***
	(0.0649)	(0.0024)	(0.0107)
Match in mobility preferences	0.0571**	0.0080***	0.1077***
	(0.0197)	(0.0013)	(0.0073)
Match in car license	0.0497***	0.0024**	−0.0677***
	(0.0019)	(0.0007)	(0.0187)
Match in motorcycle license	−0.0211	−0.0037	−0.0181
	(0.0175)	(0.0073)	(0.0903)
Match in heavy vehicle license	0.0023	−0.0005	−0.0219
	(0.0077)	(0.0030)	(0.0650)
Difference German skills	−0.0082***	−0.0002	0.0068*
	(0.0006)	(0.0002)	(0.0027)
Difference in French skills	−0.0023***	−0.0011***	−0.0104***
	(0.0005)	(0.0002)	(0.0021)
Difference English skills	−0.0023***	0.0003	0.0062**
	(0.0005)	(0.0002)	(0.0020)
Difference Italian skills	0.0007	0.0002	0.0015
	(0.0005)	(0.0002)	(0.0020)
Sufficient German skills	0.0650***	0.0033***	−0.0321***
	(0.0019)	(0.0007)	(0.0079)
Sufficient English skills	0.0424***	0.0016*	−0.0200
	(0.0023)	(0.0008)	(0.0121)
Sufficient French skills	0.0305***	0.0063***	0.0432***
	(0.0022)	(0.0007)	(0.0091)
Sufficient Italian skills	0.0319***	0.0027***	0.0014
	(0.0024)	(0.0008)	(0.0164)
Difference in workload	−0.0006***	−0.0001***	−0.0001
	(0.0000)	(0.0000)	(0.0001)
Workload match	−0.0092***	−0.0010	0.0026
	(0.0017)	(0.0006)	(0.0056)
Code Anstellungsdauer	0.0080***	0.0008*	0.0008
	(0.0012)	(0.0004)	(0.0041)
Last occupation match	0.1142***	0.0142***	0.0075
	(0.0041)	(0.0016)	(0.0042)
Last industry match	0.0618***	0.0092***	0.0152***
	(0.0025)	(0.0010)	(0.0043)
R^2	0.2916	0.0211	0.0120
Num. obs.	247230	247230	24890

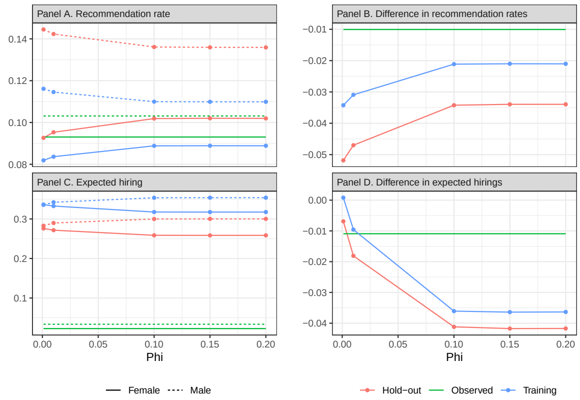
Notes: Continued from Table B.1.

Table B.2: Predicting recommendations and job matches using linear regression (continued)

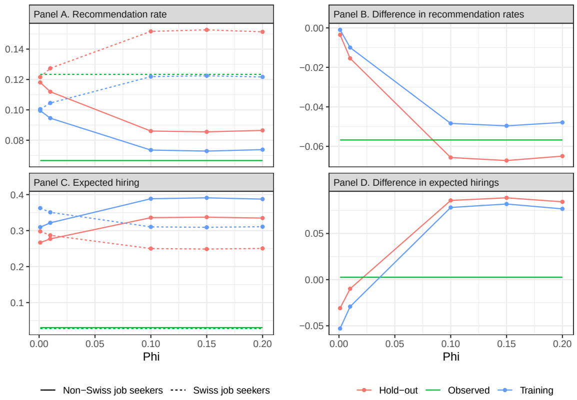
C Decongestion



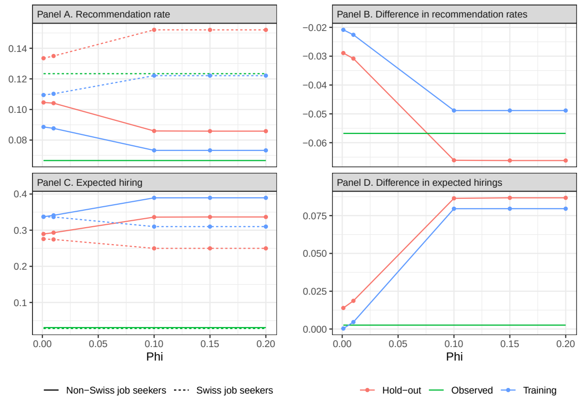
(a) DP across gender



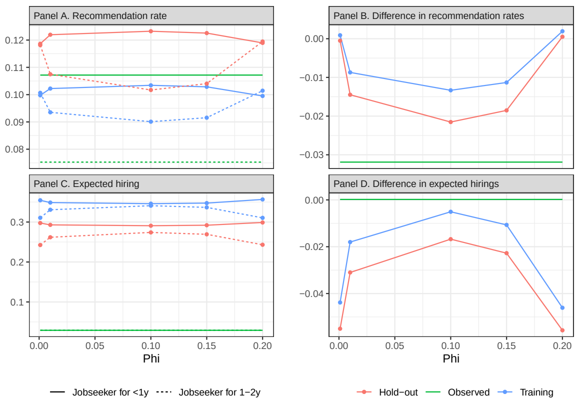
(b) PP across gender



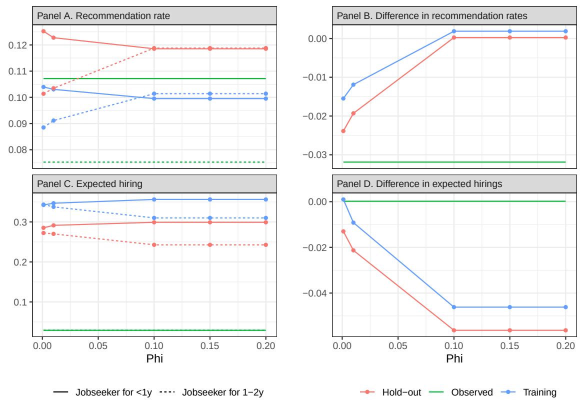
(c) DP across nationality



(d) PP across nationality



(e) DP across jobseeking duration



(f) PP across job search duration

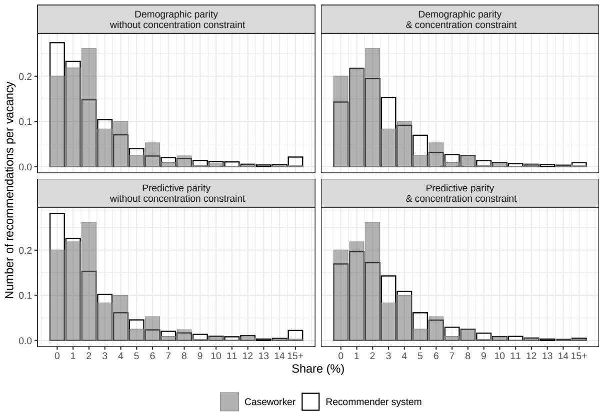
Notes: The figures report the recommendation rate (top-left panel), difference in recommendation rates (top-right panel), expected hiring rate (bottom-left panel), and gap in expected hiring rate (bottom right panel) as a function of ϕ, which determines how strictly fairness constraints are enforced. Figure C.1a and C.1b enforce demographic parity and predictive parity by gender, respectively. Figure C.1c–C.1d and Figure C.1e–C.1f focus on Swiss vs. non-Swiss nationality, and short vs. long-term job search instead. Results are reported separately for the training and hold-out sample, as well as for the observed caseworker recommendations.

Figure C.1: Recommendation share and expected hirings for different fairness-constraint recommendation systems

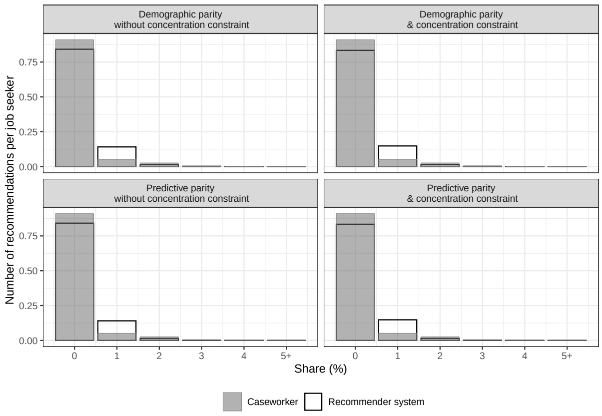
	Observed	*Recommender system*				
		Unconstr.	*Fairness constraint*		*Concentration constraint*	
Panel A. Gender						
No recommendations	0.9098	0.8417	0.8421	0.8418	0.8341	0.8341
Less than 1	0.9617	0.9826	0.9833	0.9825	0.9823	0.9821
Less than 2	0.9885	0.9974	0.9974	0.9974	0.9973	0.9973
Less than 3	0.9942	0.9993	0.9995	0.9994	0.9995	0.9993
Less than 4	0.9972	0.9999	0.9999	0.9999	0.9999	0.9999
Less than 5	0.9980	1.0000	0.9999	1.0000	0.9999	1.0000
Less than 10	0.9998	1.0000	1.0000	1.0000	1.0000	1.0000
Less than 15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Gini	0.9361	0.8584	0.8583	0.8585	0.8512	0.8514
Panel B. Nationality						
No recommendations	0.9098	0.8405	0.8429	0.8410	0.8345	0.8336
Less than 1	0.9617	0.9825	0.9827	0.9825	0.9822	0.9820
Less than 2	0.9885	0.9973	0.9975	0.9972	0.9973	0.9972
Less than 3	0.9942	0.9993	0.9993	0.9993	0.9995	0.9994
Less than 4	0.9972	0.9999	0.9999	0.9999	0.9999	0.9999
Less than 5	0.9980	0.9999	1.0000	1.0000	1.0000	1.0000
Less than 10	0.9998	1.0000	1.0000	1.0000	1.0000	1.0000
Less than 15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Gini	0.9349	0.8590	0.8597	0.8582	0.8537	0.8512
Panel C. Length of job search spell						
No recommendations	0.9079	0.8428	0.8437	0.8419	0.8372	0.8342
Less than 1	0.9608	0.9830	0.9833	0.9830	0.9828	0.9822
Less than 2	0.9883	0.9975	0.9975	0.9975	0.9975	0.9974
Less than 3	0.9940	0.9994	0.9994	0.9994	0.9995	0.9994
Less than 4	0.9970	0.9999	0.9999	0.9999	0.9999	0.9999
Less than 5	0.9979	1.0000	1.0000	1.0000	1.0000	1.0000
Less than 10	0.9998	1.0000	1.0000	1.0000	1.0000	1.0000
Less than 15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Gini	0.9349	0.8590	0.8597	0.8582	0.8537	0.8512

Notes: The table reports the share of job seekers with no or fewer than 1–15 recommendations in the observed data as well as for different recommender systems. The column 'Observed' refers to the observed data. The column 'Unconstr.' reports shares for a fairness-unconstrained recommender system. The next two columns impose demographic parity (DP) or predictive parity (PP) with $\phi = 0.001$. The last two columns add the concentration constraint in (9). The congestion is defined with respect to the distribution of the number of recommendations per vacancy.

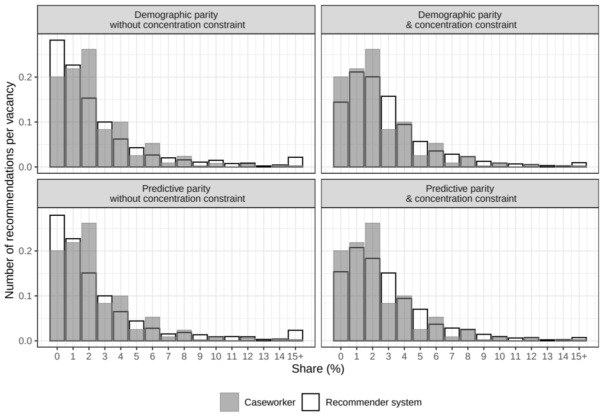
Table C.1: Number of recommendations per job seeker



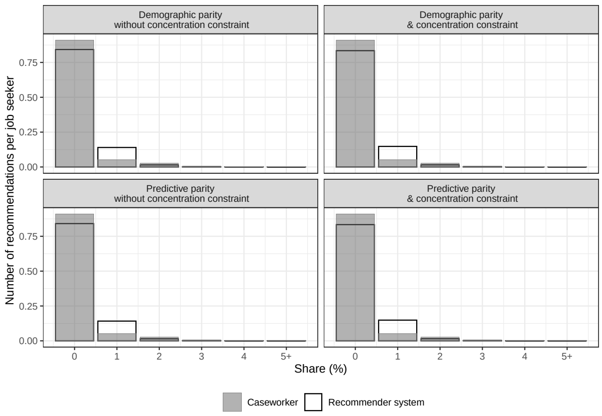
(a) Distribution across vacancies



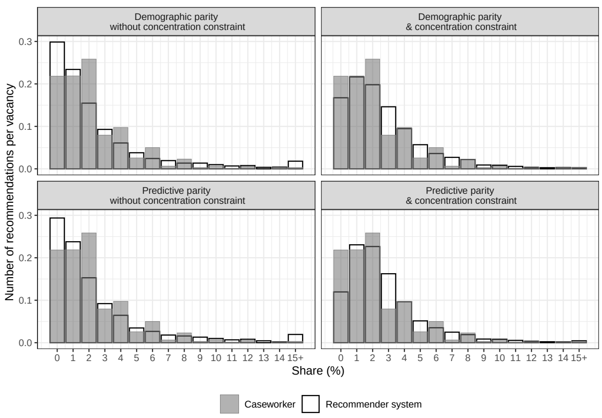
(b) Distribution across job seekers



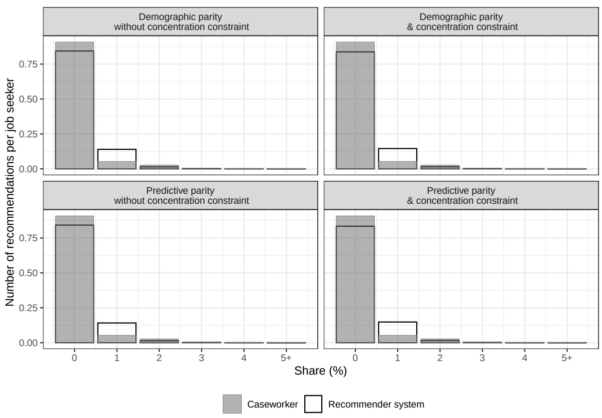
(c) Distribution across vacancies



(d) Distribution across job seekers



(e) Distribution across vacancies



(f) Distribution across job seekers

Notes: The figures show the distribution of recommendations across vacancies (left-hand side figures) and across job seekers (right-hand side figures) when enforcing demographic parity or predictive parity with and without additional concentrations constraints. Figures C.2a–C.2b enforce fairness with respect to gender. Figures C.2c–C.2d focus on nationality; Figures C.2e–C.2f on job search duration.

Figure C.2: Recommendation share and expected hirings for different fairness-constraint recommendation systems

D Estimated parameters

Constraints		*Parameter estimates*			
Fairness	*Concentration*	$\hat{\tau}_0$	$\hat{\tau}_1$	$\hat{\tau}_2$	$\hat{\tau}_3$
Panel A. Gender					
Demographic parity	No	0.0685	0.1030		
Demographic parity	Yes	0.0230	0.0513	0.0094	0.0000
Predictive parity	No	0.0917	0.0817		
Predictive parity	Yes	0.0556	0.0347	0.0028	0.0008
Panel B. Nationality					
Demographic parity	No	0.0912	0.0800		
Demographic parity	Yes	0.0398	0.0303	0.0099	0.0000
Predictive parity	No	0.0796	0.0945		
Predictive parity	Yes	0.0309	0.0502	0.0063	0.0004
Panel C. Length of job search spell					
Demographic parity	No	0.0816	0.0904		
Demographic parity	Yes	0.0299	0.0428	0.0069	0.0006
Predictive parity	No	0.0952	0.0860		
Predictive parity	Yes	0.0349	0.0232	0.0146	0.0000

Notes: The table reports the parameter estimates for $\hat{\tau}_0$, $\hat{\tau}_1$, $\hat{\tau}_2$ and $\hat{\tau}_3$ defined in (7) and (8).

Table D.1: Parameter estimates of the policy rules