A new turning point for women: artificial intelligence as a tool for reducing gender discrimination in hiring

Elena Pisanelli*

Abstract

This paper studies whether firms' adoption of artificial intelligence (AI) has a causal effect on their probability of hiring female managers. Using panel data on the 500 largest firms, measured by revenues, in Europe and the US, and a two-stage difference-in-differences I find that firms' use of AI causes, on average, an increase by 3.5% in the hiring of female managers. Exploiting heterogeneity across different types of AI I find that my result is driven by the use of assessment software, rather than that of predictive algorithms. The use of assessment software increases the share of female managers hired by companies and correlates with a reduction in firms being sued for gender discrimination in hiring. Conversely, my findings show that predictive algorithms do not affect gender inequality in managerial hires.

Keywords: artificial intelligence, gender, hiring, discrimination
JEL Codes: J71, M51

*European University Institute, Florence, IT, elena.pisanelli@eui.eu

1 Introduction

Over the past years, organizations have increased their use of artificial intelligence (AI) algorithms (Johnson et al., 2022). Their purpose is to make hiring decisions more efficient and accurate, on the basis that hiring algorithms should help mitigate recruiters' biases and ensure objectivity to the hiring process (Langenkamp, Costa, and Cheung, 2019). However, existing research shows that AI can actually discriminate based on gender and ethnicity, which casts doubt on the objectivity of artificial intelligence (Cowgill and Tucker, 2020; Gebru, 2020; Cowgill, 2019; O' Neil, 2016).

How can we explain that a technology aimed at getting rid of human bias in decision-making processes ends up being biased itself? The answer, I argue, lies in what type of AI is used. One set of AI algorithms - predictive algorithms - use the characteristics of the existing workforce to predict who should be hired (Johnson et al., 2022; Dastin, 2018; O' Neil, 2016). Another set of AI algorithms - assessment software - evaluate job applicants during the hiring process with web interviews, chats, and cognitive games to assess job candidates' skills, aptitudes, logic, and cognition (Daugherty, Wilson, and Chowdhury, 2019; Gee, 2017). In sum, AI decides who to hire based on the characteristics of firms' past workers (predictive algorithms (Rhea et al., 2022)) or based on real-time evaluations of job applicants' performance without relying on data about past workers (assessment software (Li et al., 2021)).

The existing research focuses on predictive algorithms, while leaving open questions on how assessment software may affect gender inequality in the labor market. However, predictive algorithms and assessment software can yield different effects on gender inequality in the labor market, according to how they select who to hire. On the one hand, predictive algorithms – as the existing research argues - perpetuate the existing gender inequality in the labor market because they predict who should be hired based on the characteristics of firms' past employees, thus, reproducing employers' biased hiring choices (Cowgill and Tucker, 2020; Gebru, 2020; Daugherty, Wilson, and Chowdhury, 2019; Cowgill, 2019; O' Neil, 2016). Conversely, if firms use assessment software, they may reduce gender inequality in the labor market because assessment software base their hiring choice solely on job applicants' performance evaluation and not on previous employees' data (Daugherty, Wilson, and Chowdhury, 2019; Silberg and Manyika, 2019). In this paper, I contribute to the literature by comparing and contrasting how predictive algorithms and assessment software affect gender inequality in the labor market. In particular, this paper aims at answering the following research question: do predictive algorithms and assessment software differentially affect gender inequality in the labor market? To do so, I build on statistical theory of discrimination and on expectation states theory to make predictions about the effect of assessment software and predictive algorithms on gender inequality in employment outcomes. I then empirically test my predictions using a two-stage staggered difference-in-differences approach on a panel data of the 500 largest companies, measured by revenues, in Europe and the US over 8 years (2013-2021). To measure gender inequality in employment outcomes I consider gender inequality in companies' managerial pools, since gender inequality in leadership positions explains most of the persistence in the gender wage gap (Mandell et al. 2022).

I find that firms' use of AI causes, on average, an increase by 3.5% in the hiring of female managers. Exploiting heterogeneity across different types of AI, I find that my result is driven by the use of assessment software, rather than that of predictive algorithms. The use of assessment software increases the share of female managers hired by companies and correlates with a reduction in firms being sued for gender discrimination in hiring. Conversely, my findings show that predictive algorithms do not affect gender inequality in managerial hires.

2 Theoretical framework

When firms rely on AI algorithms in hiring, they can use assessment software and predictive algorithms. The distinction between the two is of crucial importance: predictive algorithms utilize performance history of previous employees, and assessment software does not.

To understand how such AI algorithms can affect gender discrimination in hiring end gender inequality in employment outcomes, I build on statistical theory of discrimination and on expectation states theory. I consider these theories because they are among the most influential theories in economics and sociology in explaining how and when employers may discriminate based on gender.

Arrow (1973)'s and Phelps (1972)'s statistical theory of discrimination postulates that employers have imperfect information about the future productivity of job candidates, which gives them an incentive to use easily observable ascriptive characteristics, such as race or gender, to infer the expected productivity of applicants (Correll and Benard, 2006). The theory, thus, highlights a key explanation for why employers discriminate based on gender: the excessive cost of gaining information about individual workers' productivity leads employers to rely on their biased beliefs about group statistics to infer workers' individual productivity (Arrow, 1973; Phelps, 1972). Discrimination, thus, arises as a rational solution to an information problem.

Expectation states theory, as with statistical theory, assumes that employers discriminate against women because they have biased beliefs about workers' competence, which give rise to biased expectations about workers' future performance if hired. However, biased "status beliefs" do not arise from biased statistical evidence but are defined as "widely held cultural beliefs that link greater social significance and general competence, as well as specific positive and negative skills, with one category of a social distinction (e.g., men) compared to another" (Ridgeway, 2001, p. 638). These beliefs affect gender inequality in the labor market through both demand and supply side mechanisms. This paper focuses on the demand side, where the existing sex-segregated structure of the contemporary labor market makes cultural beliefs about gender salient (Ridgeway, 2011). The gender stereotype pertaining to the sex that predominates in a job biases the traits of the ideal worker for that job (Reskin, 2005). For example, being a nurse is typically a female job and is associated to female traits (e.g., empathy, compassion, caregiving, ...), leading recruiters to prefer women over men for such a job. Further, the prestige associated with the job itself can make cultural beliefs about gender salient (for a thorough discussion see (Ridgeway, 2019)). In fact, jobs associated with authority and competence are usually culturally associated with masculinity (Powell, 2002). Cultural beliefs about gender, thus, set the stage for gender discrimination in hiring from the demand side by biasing employers' perceptions of men and women's competence to perform a given job (Ridgeway, 2011; Heilman and Okimoto, 2007; Gorman, 2005). Further and more problematically, when women apply for authority positions or self-present as agentic and assertive, they are punished for it (Quadlin, 2018; Heilman and Okimoto, 2007; Rudman and Glick, 2001). This is because by self-presenting themselves as agentic, incongruity arises between women's behavior and cultural beliefs that presume lower status position and communality for women (Rudman and Glick, 2001; Prentice and Carranza, 2002). This incongruity leads to a backlash on the demand side, with employers punishing agentic women.

How do discrimination theories relate to predictive algorithms and assessment software? Predictive algorithms use firms' past workers' productivity and characteristics as a benchmark for evaluating job applicants' performance (Rhea et al., 2022). In the context of the statistical theory of discrimination, predictive algorithms use the average information about the group to which the job applicant belongs to infer how the job applicant will perform if hired. More precisely, predictive algorithms hitherto have been programmed to estimate the average group (e.g., men, women) probability of success for firms' past workers, to then predict the future performance of individual job applicants belonging to the same group (Langenkamp, Costa, and Cheung, 2019). Thus, how predictive algorithms work is not any different from how human recruiters make their hiring decisions, they both use surrogate and biased information about the group to infer the individual job applicants' productivity.

In the context of expectation states theory, predictive algorithms are not subject to status beliefs, because, after all, they are a software. However, relying on data about firms' past employees, predictive algorithms associate gender to the probability of success in the job. If employers hired mostly men, predictive algorithms carry over employers' biased hiring decisions and biased beliefs (Langenkamp, Costa, and Cheung, 2019). My first prediction, thus, is:

Hypothesis 1 *firms' use of predictive algorithms in hiring reproduces the existing gender inequality (equality) in firms' managerial pools.*

This hypothesis is in line with the existing research showing that AI makes biased hiring decisions because it learns to be biased from humans (Gonzalez et al., 2022; Black and van Esch, 2020; Gebru, 2020; Kochling and Wehner, 2020; Daugherty, Wilson, and Chowdhury, 2019; Silberg and Manyika, 2019; Bogen, 2019; O' Neil, 2016).

Consider now the assessment software, which evaluates job applicants based on their performance in cognitive tests, interviews, and chats (Li et al., 2021). In the context of the statistical theory of discrimination, assessment software gains vast amounts of information about individual job applicants' productivity and uses such information to decide who should be hired (Daugherty, Wilson, and Chowdhury, 2019). Therefore, assessment software does not rely on the biased information about the group to infer individual job applicants' productivity but provides employers with accurate information on individual job applicants' productivity. This key difference from predictive algorithms should give assessment software no statistical reason to discriminate based on gender.

In the context of expectation states theory, as with predictive algorithms, assessment software is not subject to status beliefs when evaluating job applicants because it is a software and not a human being. Further, because assessment software evaluates job applicants without using past employees' data, it should not carry over employers' biased beliefs and biased hiring decisions. My second prediction, thus, is:

Hypothesis 2 *firms' use of assessment software in hiring decreases gender inequality in firms' managerial pools.*

3 Data and Method

3.1 Data

I perform the analyses in this paper on those European and American firms that entered Fortune Global 500 in 2021. This dataset includes European and American companies ranked as the top 500 corporations worldwide as measured by revenue. The main reason to focus on such specific firms is that I can collect public available information on whether they use AI in hiring or not. Further, these firms have financial and organizational similarity. Namely, they have a very similar productivity and size (see Table 1), which affect their innovative capabilities (Gòmez and Vargas, 2012) and, thus, their potential adoption of AI in hiring. The public availability of data on the use of AI and the similarity among companies allow me to study the effect of firms' use of AI on their share of female managers hired. To do so, I match firms that use AI with the most similar firms that do not use it. Information on firms' characteristics come from the Orbis database by Bureau van Dijk.

The solution developed, [redacted], uses AI to leverage information about the job market and past experiences of employees to predict the candidates most likely to be successful.

(a) Predictive algorithms

The software derives required skills from job requisitions and generates a match score against skills described in resumes. These scores predict future job performance.

(b) Assessment software

Figure 1: Example of AI use in firms' annual reports

I collect data on the variables that can determine differences in firms' adoption of a new technology, that is, where each firm's headquarters are based, the industry in which the firm operates, firm's productivity, total assets, return on equity, profit margin, net income, number of employees (Antonelli, Orsatti and Pialli, 2022; Gòmez and Vargas, 2012), and share of female directors to control for gender diversity in firms' leadership roles. Note that the share of female directors differs from the share of female managers, because directors are nominated by firms' shareholders while managers are hired. Therefore, AI would not play any direct role in shaping gender (in)equality in the board of directors. I collect data for the years going from 2013 to 2021.

My dependent variable is firms' share of female managers. The variable represents the yearly number of female managers hired in every firm over the total managers hired for that year. Data on firms' use of AI in hiring are, instead, much harder to get. In order to know whether firms use predictive algorithms or assessment software in hiring and when they have adopted it, I manually extracted information on firms' use of AI from firms' publicly available annual integrated reports. Figure 1 shows an example of information about AI use in hiring on firms' annual integrated report. Note that in my sample, firms never mention the use of both predictive algorithms and assessment software in hiring but either one or the other.

I assign each firm to the assessment software treatment (Da=1) if (1) its annual integrated report contains evidence of the use of assessment software in hiring, and/or (2) there is evidence on the firm's website or online that the firm uses AI-powered logic, aptitude and reasoning tests, video interviews and chats to evaluate job candidates (Li et al., 2021). I assign each firm to the predictive algorithms treatment (Dp=1) if (1) its annual integrated report contains evidence of the use of predictive algorithms in hiring, and/or (2) there is evidence on the firm's website or online that the firm uses predictive algorithms to evaluate job candidates (Rhea et al., 2022).

3.2 Empirical strategy

The empirical strategy relies on a two-stage staggered difference-in-difference with ex-ante matching. I do an ex-ante matching because my control group exceeds in number my treatment group.

The identification strategy, thus, is developed in two stages: (i) I match, based on observable characteristics, treated and control firms using Mahalanobis distance matching (Mahalanobis, 1936); (ii) I perform a two-stage staggered difference-in-differences estimation with multiple time periods, relying on Butts and Gardner (2021) because not all treated firms in my dataset have adopted AI in hiring in the same year (staggered difference-in-differences) and the choice to adopt AI in hiring is endogenous (instrumental variable approach).

3.2.1 Mahalanobis distance matching

The first step of the empirical strategy resorts to using Mahalanobis distance matching to balance the treated and control firms on observable covariates. Relying on Mahalonobis distance matching before the difference-in-differences estimate allows to account for the systematic dynamic differences between those firms which use AI in hiring and those which do not.

For each treated firm, with treatment D_i defined as the use of AI in hiring ($D_i = 1$), I find all available untreated firms ($D_i = 0$) with the most similar — in terms of Mahalanobis distance — variables $\{x\}$ that may determine differences in firms' adoption of AI in hiring (firms' size, firms' productivity, profit margin, return on equity, industry, and country). Equation 1 presents the econometric specification of the Mahalanobis distance definition.

$$d(u, v) = (u - v)^T C_{OR}^{-1}(u - v) \tag{1}$$

With u and v values of $\{x^T, \hat{q}(x)\}^T$, where x are the observable covariates and $\hat{q}(x)$ is the estimated log odds against exposure to treatment; and C_{OR} sample covariance matrix of $\{x^T, \hat{q}(x)\}$ in the control group (Rosenbaum and Rubin, 1985).

The Appendix reports the distribution of the covariates used for matching treated and control firms with Mahalanobis distance matching (for each year between 2013 and 2021). Existing research shows that the probability that firms introduce innovation in their processes or products significantly depends on firms' size (Gòmez and Vargas, 2012). I, therefore, match treated and control firms — before estimating through staggered difference-in-differences the effect of AI on firms' probability to hire female managers — on measures of firms' size, that is, total assets and number of employees (Damanpour, 1992). Further, the probability that firms adopt innovations both within their processes and products is strongly associated with firms' profits and financial resources (Antonelli, Orsatti and Pialli, 2022). I, therefore, match treated and control firms also on productivity, profit margin and return on equity. Last, since existing studies show that the relationship between innovation and the above variables is not the same across countries and industries Damanpour, 1992), I match treated and control firms on industry (4 digits NACE code) and country. In particular, the Appendix presents the balancing achieved after the Mahalanobis distance matching. Table 1 reports descriptive statistics of all variables over the full sample and the restricted matched sample.

Table 1: Summary statistics

	Full sample		Restricted matched sample		T-test
	Mean	SD	Mean	SD	
Share of female managers hired	0.026	0.05	0.024	0.042	NS
Use of AI	0.107	0.309	0.298	0.458	*
Log assets	25.262	1.469	25.748	1.368	*
Log productivity	24.406	0.824	24.349	0.528	NS
Net income (millions)	12.357	18.577	12.629	13.382	NS
Profit margin	13.286	14.108	15.353	11.996	*
Return on equity	27.986	67.858	16.954	12.365	*
Log employees	11.278	1.115	11.489	0.63	*
Share of female directors	0.052	0.044	0.053	0.039	NS
N	1,621		531		

Note that the two groups do not differ in the distribution of the dependent variable (share of female managers). The restricted sample has a slightly higher share of firms using AI.

3.2.2 Two-stage staggered difference-in-differences on matched firms

I estimate, for matched firms, the effect of using AI on firms' share of female managers hired through a two-stage staggered difference-in-differences technique with multiple time periods, relying on Butts and Gardner (2021)'s approach.

The first stage of the procedure consists of a regression of outcomes on group and period fixed effects, estimated using the subsample of untreated observations. In the second stage, the estimated group and period effects are subtracted from observed outcomes, and these adjusted outcomes are regressed on treatment status. Under the usual parallel trends assumption, this procedure identifies the overall average effect of the treatment on the treated (i.e., across groups and periods), even when average treatment effects are heterogeneous over groups and periods (Butts and Gardner, 2021, p.4).

The two-stage estimation procedure first estimates:

$$Y_{gpit} = \lambda_g + \gamma_p + \epsilon_{gpit} \tag{2}$$

for non-treated units. With g treatment group, p period, i unit, and t time variable, λ_g group fixed effect, γ_p time fixed effect. From equation 2, the estimated group fixed effect $(\hat{\lambda_g})$ and time fixed effect $(\hat{\gamma_p})$ are retained.

Second, the adjusted outcomes $Y_{gpit} - \hat{\lambda_g} - \hat{\gamma_p}$ are regressed on the treatment variable. Conditional on the parallel trend assumption, the procedure identifies $E(\beta_{gp}|D_{gp} = 1)$ even when the the adoption and average effects of the treatment are heterogenous with respect to groups and periods (Butts and Gardner, 2021). In particular, the expected treatment effect is estimated as group×period-specific average treatment effects (Butts and Gardner, 2021):

$$E(\beta_{gp}|D_{gp} = 1) = \sum_{g=1}^{G} \sum_{p=g}^{P} \beta_{gp} P(g, p|D_{gp=1}) \tag{3}$$

When estimating $E(\beta_{gp}|D_{gp} = 1)$, I weight each observation by the weight generated with Mahalanobis distance matching, in order to condition the estimate of the weighted average treatment effect on the observable covariates.

The event study setting develops by regressing in the second stage the adjusted outcomes $Y_{gpit} - \hat{\lambda_g} - \hat{\gamma_p}$ on $D_{Rgp}, ..., D_{0gt}, ...D_{Pgp}$, with R periods of treatment.

As with staggered difference-in-difference, in order for the estimated ATT to be valid and reliable, a series of assumptions should be imposed (Callaway and Sant'Anna, 2021).

Assumption 1: Limited treatment anticipation. The assumption states that firms should not anticipate treatment by any period. The assumption is very likely to hold in the context of this paper, since it is unlikely that firms increase the hiring of female managers in sight of AI adoption. However, I account for potential anticipation of the treatment by allowing for anticipatory behavior and imposing conditional parallel trends in pre-treatment periods, making the parallel trend assumption (discussed in the next paragraph) stronger.

Assumption 2: Conditional parallel trends based on a never-treated group. The assumption imposes that, conditional on covariates, the average outcomes for the firms first treated in group g (with g year of AI adoption) and for the never-treated firms would have followed parallel paths in the absence of the treatment. Section 4 provides evidence of the validity of the conditional parallel trend assumption.

What makes firms' use of AI in hiring endogenous is the fact that firms may use AI under the premise that it should help mitigate recruiters' biases and ensure objectivity to the hiring process (Langenkamp, Costa, and Cheung, 2019). Therefore, firms that have more developed ESG strategies or that lack diversity and inclusion in their workforce may be more likely to use AI for improving diversity in their hiring outcomes. Further, my estimates may suffer from simultaneity bias. Therefore, I instrument firms' use of AI in hiring with firms' intangible fixed assets lagged at time t-1, retrieved from the Orbis database by Bureau Van Dijk. Intangible fixed assets are formally defined as "all intangible assets such as formation expenses, research expenses, goodwill, development expenses and all other expenses with a long-term effect" (Altomonte et al., 2022, p.5). Firms' intangible fixed assets are typically used in the existing economics literature as a proxy for firms' use of AI (see, for example, Agarwall et al. (2021) or Corrado et al. (2021)). AI might, in fact, be thought of as spending on software and databases (Corrado et al., 2021). Firms' intangible fixed assets comprise all expenses that a firm makes to form employees, to undertake research and development, and to innovate. It seems, thus, reasonable to assume that firms' intangible fixed assets should not directly affect the share of female managers hired, except through AI. The lag at time t-1 allows me to address potential simultaneity bias.

4 Results

4.1 Effect of instrumented AI on firms' share of female managers hired

This section presents the estimated effect of the instrumented AI on the share of female managers hired. Figure 2 shows the graphical representation of the event study. Table 2 reports the simple weighted average estimate of the ATT. Standard errors are clustered at firm level.

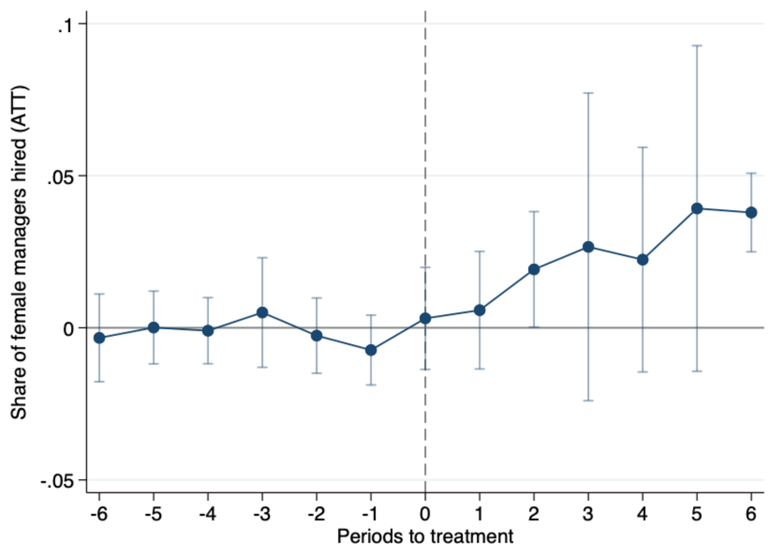


Figure 2: Two-stage staggered difference-in-differences results. Event study

Table 2: Two-stage staggered difference-in-differences results.

Panel a. Reduced form (simple weighted average ATT)	
	Share of female managers hired
	(1)
Firms' use of AI	0.0354*
	(2.14)
Managers' age	✓
Total assets	✓
Number of employees	✓
Productivity	✓
Profit margin	✓
Net income	✓
Country	✓
Industry	✓
Return on equity	✓
Share of female directors	✓
Obs.	1,467
Mean of the share of female managers hired	0.024
Panel b. First stage	
	Firms' use of AI
	(1)
Lagged intangible fixed assets	0.019***
	(0.003)
Managers' age	✓
Total assets	✓
Number of employees	✓
Productivity	✓
Profit margin	✓
Net income	✓
Country	✓
Industry	✓
Return on equity	✓
Share of female directors	✓
Year fe	✓
Unit fe	✓
F	20.152
Obs.	1,467

Standard errors in parentheses, clustered at firm level
*** p<0.01, ** p<0.05, * p<0.1

As Figure 2 shows, the conditional parallel trend assumption is satisfied, since the estimates of the ATT in each year before treatment (t-6 to t0) are not statistically different from zero.

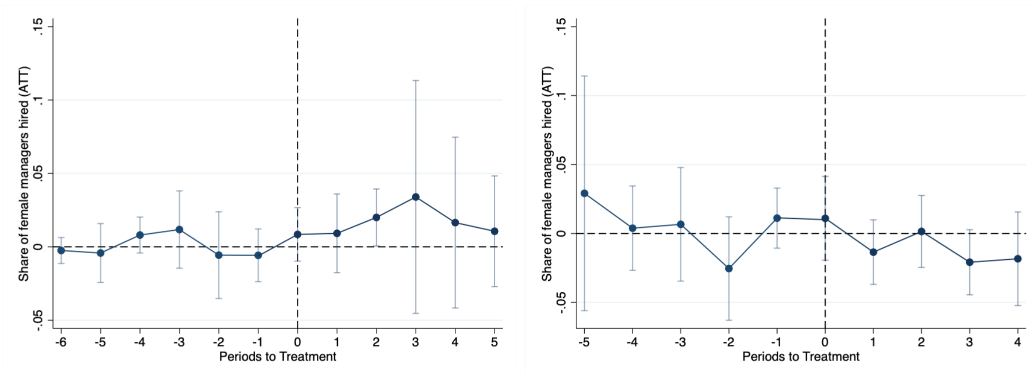
The estimated simple weighted average ATT in Table 2 shows that firms' use of AI in hiring causes, on average, an increase by 3.5% in the share of female managers hired. This effect is considerable in magnitude, since – as reported in Table 2 and in Table 1 – the average share of female managers hired in the sample is only 2%.

The estimated results suggest that firms' use of AI can help reducing the persistent under-representation of women in managerial positions. An important implication is that by increasing female representation in firms' managerial positions, firms' use of AI can promote the overall gender equality in the labor market. This because the presence of women in managerial positions may encourage female participation in the workforce by establishing a role model (Porter and Serra, 2020). An increase of female participation in the workforce can improve fertility and households' well-being (Cohen and Huffman, 2007; Hensvik, 2014; Profeta, 2020).

Even if my result shows that AI can increase the share of women hired in managerial positions, we still do not know what type of AI drives this effect. The next section shows the decomposition of the result by type of AI. The analyses that exploit the heterogeneity in the type of AI used by firms rely on the staggered difference-in-difference technique proposed by Callaway and Sant'Anna (2021). I do not instrument the type of AI firms can use because it seems reasonable to assume that the endogeneity problem arises at the extensive margin, *i.e.*, when firms decide whether to use AI in hiring or not, rather than at the intensive margin - the type of AI firms adopt. Therefore, conditional on firms using AI, I assume the choice of which type of AI to use is exogenous.

4.2 Decomposition of the effect by type of AI

Figure 3 shows the graphical representation of the event study. Table 3 reports the simple weighted average estimate of the ATT. Standard errors are clustered at firm level.



(a) Assessment software

(b) Predictive algorithms

Figure 3: Staggered difference-in-differences results by type of AI. Event study

Table 3: Staggered difference-in-differences results by type of AI

	Share of female managers hired (simple weighted average ATT) (1)	(2)
Firms' use of assessment software	0.016*	
	(0.008)	
Firms' use of predictive algorithms		-0.005
		(0.007)
Managers' age	✓	✓
Total assets	✓	✓
Number of employees	✓	✓
Productivity	✓	✓
Profit margin	✓	✓
Net income	✓	✓
Country	✓	✓
Industry	✓	✓
Return on equity	✓	✓
Share of female directors	✓	✓
Innovative activity	✓	✓
Obs.	1,458	1,458

Standard errors in parentheses, clustered at firm level
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

As Figure 3 and Table 3 show, the increase in the probability of hiring female managers is driven by firms' use of assessment software; the effect of firms' use of predictive algorithms is not statistically different from zero.

My results provide evidence in favor of assessment software reducing gender inequality in firms' managerial pools (hypothesis 2). They also provide evidence in favor of predictive algorithms not affecting gender inequality in firms' managerial pools. Therefore, it seems reasonable to argue that predictive algorithms reproduce the existing gender inequality (equality) in firms' managerial pools (hypothesis 1).

4.3 Mechanism: reduction of gender discrimination in hiring

The main finding that firms' use of assessment software increases the share of female managers hired may be explained by the likelihood of assessment software decreasing gender discrimination in hiring. I, thus, estimate using staggered difference-in-differences (Callaway and Sant'Anna, 2021) the effect of firms' use of assessment software and gender discrimination in hiring. To measure gender discrimination in hiring, I collected publicly available documentation on the court cases in which firms were sued for gender discrimination in hiring between 2013 and 2021. Because of public availability of lawsuits and the relative documentation, I was able to collect information only on firms in the US. This means that in the following analyses, my results do not apply anymore to both European and US firms. However, as Figure 4 shows, performing my analyses only on US firms does not affect the reliability of my results. I can, thus, restrict the sample to only US firms, amounting to 56% of my original sample.

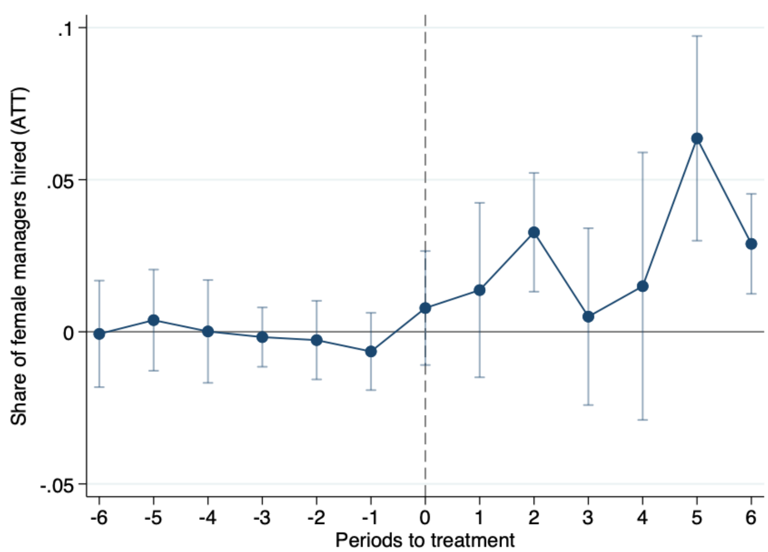


Figure 4: Two-stage staggered difference-in-differences results. Event study on US firms

Figure 5 shows the graphical representation of the event study. Table 4 reports the simple weighted average estimates of the effect of using assessment software on firms' probability of being sued for gender discrimination in hiring.

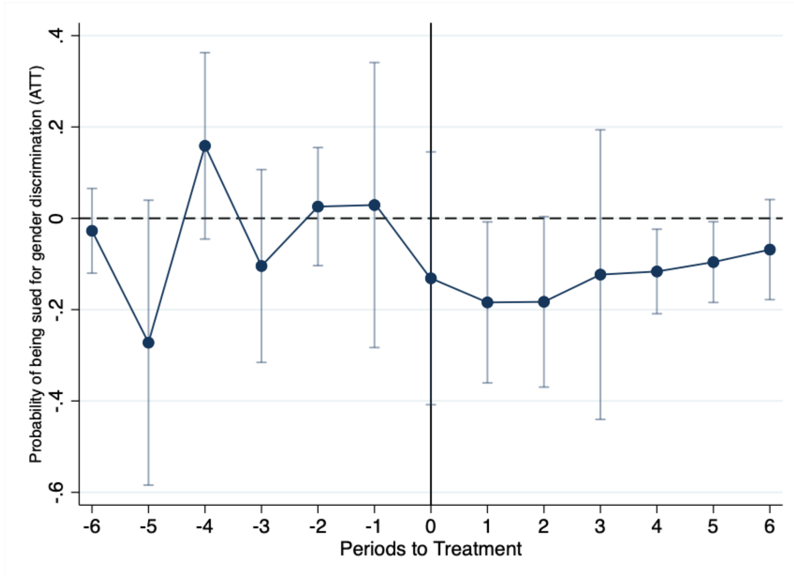


Figure 5: Staggered difference-in-differences results. Event study. US firms' lawsuits

Table 4: Staggered difference-in-differences results. US firms' lawsuits.

	Probability of being sued for gender discrimination in hiring (simple weighted average ATT)
	(1)
Firms' use of assessment software	-0.155*
	(0.089)
Managers' age	✓
Total assets	✓
Number of employees	✓
Productivity	✓
Profit margin	✓
Net income	✓
Country	✓
Industry	✓
Return on equity	✓
Share of female directors	✓
Innovative activity	✓
Obs.	819

Standard errors in parentheses, clustered at firm level
*** p<0.01, ** p<0.05, * p<0.1

Using assessment software coincides with a reduction by 15% in the probability of firms being sued for gender discrimination in hiring. This result hints at the possibility that using assessment software helps firms to decrease gender discrimination in hiring, which in turn results in an increase in the share of female managers hired.

5 Conclusion

This study used a two-stage staggered difference-in-differences approach to test whether firms' use of AI may affect their hiring of female managers.

As discussed in the introduction, firms are increasingly using AI algorithms in hiring.

Their purpose is to make hiring decisions more efficient and accurate, on the basis that hiring algorithms should help mitigate recruiters' biases and ensure objectivity to the hiring process. However, existing research shows that AI can actually discriminate based on gender and ethnicity, which casts doubt on its objectivity. This paper proposes that the explanation lies in the type of AI firms use in hiring. I argue in this paper that while predictive algorithms reproduce the existing gender inequality (equality), assessment software reduce gender inequality in firms' managerial pools. The fact that predictive algorithms decide who should be hired based on the characteristics of previous employees leads them to inherit human biases. However, when firms use assessment software in hiring, they evaluate job applicants based on their individual performance without using past workers' data. This leads firms to reduce gender inequality in managerial hires. I find AI increases by 3.5% firms' share of female managers hired. This result is driven by assessment software, while the effect of predictive algorithms cannot be claimed statistically different from zero. Further, I show assessment software are correlated with a reduction in firms being sued for gender discrimination in hiring.

My results answer to the central puzzle around AI by suggesting that what is discussed in the literature as evidence of AI being biased can only be circumscribed to predictive algorithms. Conversely, assessment software reduces gender inequality in the labor market. This paper adds to the lack of micro-level evidence on the impact of firms' use of AI on the persistent female under-representation in managerial positions. It is also the first quasi-experimental evaluation of AI on key employment outcomes.

Acknowledgments

The author gratefully acknowledges Klarita Gërxhani; Giovanni Mastrobuoni; Federica Maria Origo; Arnout van de Rijt; Emilio J. Castilla for the insights and precious feedback. William Foley; Amanda Haraldsson; Mena Lüsse; Jonathan Winter; Guido Pialli; Luca Favero for the fruitful discussions; the participants to the 2022 EEA - ESEM annual congress.

References

Akerlof, G.A. (1997) "Social Distance and Social Decisions". Econometrica. 65(5): 1005–1027.

Albert, E.T. (2019) "AI in talent acquisition: a review of AI applications used in recruitment and selection". Strategic HR Review. 18(5): 215–221.

Altomonte, C., Bauer, P., Gilardi, A. M., Soriolo, C. (2022). Intangible assets, industry performance and finance during crises. BAFFI CAREFIN Centre Research Paper, (2022-173).

Arrow, K.J. (1973) "The Theory of Discrimination". O. Ashenfelter, A. Rees, eds. Discrimination in Labor Markets (Princeton University Press, Princeton), 3–33.

Becker, G. S. (1957). The economics of discrimination. University of Chicago press.

Bertrand, M., A. Schoar (2003) "Managing with style: The effect of managers on firm policies". The Quarterly journal of economics. 118(4): 1169–1208.

Bildfell, C. (2019) "Hiring Algorithms in the Canadian Private Sector: Examining the Promise of Greater Workplace Equality". Canadian Journal of Law and Technology. 17(2): 115.

Black, J. S., P. van Esch (2020). "AI-enabled recruiting: What is it and how should a manager use it?". Business Horizons. 63(2): 215 – 226.

Bogen, M. (2019). "All the ways hiring algorithms can introduce bias". Harvard Business Review. 6, 2019.

Brathwaite, T., A. Vij, J. L. Walker. (2017) “Machine learning meets microeconomics: The case of decision trees and discrete choice”. arXiv preprint. 1711(04826).

Breines, W. (2006) “The trouble between us: An uneasy history of white and black women in the feminist movement” (Oxford University Press, New York).

Brescoll, V. L. (2016) “Leading with their hearts? How gender stereotypes of emotion lead to biased evaluations of female leaders”. The Leadership Quarterly. 27(3): 415–428.

Brescoll, V. L., T. G. Okimoto, A. C. Vial (2018) ”You’ve come a long way. . . maybe: How moral emotions trigger backlash against women leaders”. Journal of Social Issues. 74(1): 144–164.

Butts, K., Gardner, J. (2021). did2s: Two-stage difference-in-differences. arXiv preprint arXiv:2109.05913.

Callaway, B., P. H. Sant’Anna (2021) “Difference-in-Differences with multiple time periods”. Journal of Econometrics. 225(2): 200–230.

Castilla, E. J. (2008) “Gender, race, and meritocracy in organizational careers”. American Journal of Sociology. 113:1479–1526.

Castilla, E. J. (2015) “Accounting for the Gap: A Firm Study Manipulating organizational Accountability and Transparency in Pay Decisions”. Organization Science. 26(2):311-333.

Cerulli, G. (2019) “Data-driven sensitivity analysis for matching estimators”. Economics Letters. 185(2019): 108749.

Cohen, P. N., M. L. Huffman (2007) “Working for the woman? Female managers and the gender wage gap”. American sociological review. 72(5): 681–704.

Corrado, C., Haskel, J., Jona-Lasinio, C. (2021). Artificial intelligence and productivity: An intangible assets approach. Oxford Review of Economic Policy, 37(3), 435-458.

Correll, S. J. (2001). "Gender and the career choice process: The role of biased self-assessments". American journal of Sociology, 106(6), 1691-1730.

Correll, S. J., S. Benard (2006) "Biased estimators? Comparing status and statistical theories of gender discrimination". S.R. Thye, E.J. Lawler, eds. Advances in group processes (Emerald Group Publishing Limited), 89–116.

Correll, S. J., Benard, S., Paik, I. (2007). "Getting a job: Is there a motherhood penalty?". American journal of sociology, 112(5), 1297-1338.

Cowgill, B. (2019) "Bias and productivity in humans and machines". Working paper, Columbia Business School.

Cowgill, B., C. E. Tucker (2020) "Algorithmic fairness and economics". Working paper, Columbia Business School.

Crittenden, A. (2001) "The Price of Motherhood: Why the Most Important Job in the World Is Still the Least Valued" (Metropolitan Books, New York).

Cuddy, A., S. T. Fiske, P. Glick (2004) "When professionals become mothers, warmth doesn't cut the ice". Journal of Social Issues. 60(4): 701–718.

Cunningham, J., T. Macan (2007) "Effects of applicant pregnancy on hiring decisions and interview ratings". Sex Roles. 57(7): 497–508.

Damanpour, F. (1992) "Organizational size and innovation". firm studies, 13(3): 375–402.

Dastin, J. (2018). "Amazon scraps secret AI recruiting tool that showed bias against women". In Ethics of Data and Analytics (pp. 296-299). Auerbach Publications.

Datta, A., M. C. Tschantz, A. Datta (2015) "Automated experiments on ad privacy settings". Proceedings on Privacy Enhancing Technologies. 2015(1): 92–112.

Daugherty, P. R., H. J. Wilson, R. Chowdhury (2019) "Using artificial intelligence to promote diversity". MIT Sloan Management Review. 60(2).

DeBrusk, C. (2018) “The risk of machine-learning bias (and how to prevent it)”. MIT Sloan Management Review (online). https://sloanreview.mit.edu/article/the-risk-of-machine- learning-bias-and-how-to-prevent-it/

Dezs¨o, C. L., D. G., Ross (2012) “Does female representation in top management improve firm performance? A panel data investigation”. Strategic management journal, 33(9): 1072–1089.

Eagly, A. H. (1997) “Sex differences in social behavior: Comparing social role theory and evolutionary psychology”. American Psychologist. 52(12): 1380–1383.

Fernandez, R. M., M. L. Mors (2008) “Competing for jobs: Labor queues and gender sorting in the hiring process”. Social Science Research. 37(4): 1061–1080.

Fiske, S. T., Cuddy, A. J., Glick, P., Xu, J. (2002). “A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition”. Journal of Personality and Social Psychology, 82(6), 878-902.

Frankish, K., W. M. Ramsey (2014) “The Cambridge Handbook of Artificial Intelligence” (Cambridge University Press, Cambridge).

Gebru, T. (2020) “Race and gender”. M. D. Dubber, F. Pasquale, S. Das, eds. The Oxford handbook of ethics of AI. (Oxford University Press, New York), 251–269.

Gee, K. (2017). In Unilever’s radical hiring experiment, resumes are out, algorithms are in. Wall Street Journal, 26.

Gloor, J. L., M. Morf, S. Paustian-Underdahl, U. Backes-Gellner (2020) “Fix the Game, Not the Dame: Restoring Equity in Leadership Evaluations”. J ournal of Business Ethics. 161(3): 497–511.

Goldin, C., C. Rouse (2000) “Orchestrating impartiality: The impact of” blind” auditions on female musicians”. American economic review. 90(4): 715–741.

Gonzalez, M. F., Liu, W., Shirase, L., Tomczak, D. L., Lobbe, C. E., Justenhoven, R., Martin, N. R. (2022). “Allying with AI? Reactions toward human-based, AI/ML-based, and augmented hiring processes”. Computers in Human Behavior, 130, 107179.

Hensvik, L. E. (2014) “Manager Impartiality: Worker-Firm Matching and the Gender Wage Gap”. ILR Review. 67(2):395–421.

Hewlett, S.A., K. Peraino, L. Sherbin, K. Sumberg (2010) “The sponsor effect: Breaking through the last glass ceiling”. Harvard Business Review.

Hunkenschroer, A. L., Luetge, C. (2022). “Ethics of AI-enabled recruiting and selection: A review and research agenda”. Journal of Business Ethics, 1-31.

Johnson, B. A., Coggburn, J. D., Llorens, J. J. (2022). “Artificial Intelligence and Public Human Resource Management: Questions for Research and Practice”. Public Personnel Management, 00910260221126498.

Kahneman, D., S. P. Slovic, A. Tversky (1982) “Judgment under uncertainty: heuristics and biases” (Cambridge University Press, Cambridge).

K¨ochling, A., Wehner, M. C. (2020). “Discriminated by an algorithm: a systematic review of discrimination and fairness by algorithmic decision-making in the context of HR recruitment and HR development”. Business Research, 13(3), 795-848.

Lambrecht, A., C. E. Tucker (2019) “Algorithmic Bias? An Empirical Study into Apparent Gender-Based Discrimination in the Display of STEM Career Ads”. Management Science. 65(7): 2966–2981.

Langenkamp, M., A. Costa, C. Cheung (2019) “Hiring Fairly in the Age of Algorithms”. Working paper. https://ssrn.com/abstract=3723046.

Iyer, R., Li, Y., Li, H., Lewis, M., Sundar, R., Sycara, K. (2018). “Transparency and explanation in deep reinforcement learning neural networks”. In Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society (pp. 144-150).

Li, D., Raymond, L. R., Bergman, P. (2020). "Hiring as exploration" (No. w27736). National Bureau of Economic Research.

Li, L., Lassiter, T., Oh, J., Lee, M. K. (2021, July). "Algorithmic hiring in practice: Recruiter and HR Professional's perspectives on AI use in hiring". In Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society (pp. 166-176).

Mahalanobis, P. C. (1936). "On the generalized distance in statistics". National Institute of Science of India.

Newman, D. T., N. J. Fast, D. J. Harmon (2020) "When eliminating bias isn't fair: Algorithmic reductionism and procedural justice in human resource decisions". Organizational Behavior and Human Decision Processes. 160: 149–167.

Niederle, M., L., Vesterlund (2011) "Gender and competition". Annual Review of Economics, 3(1): 601–630.

O' Neil, C. (2016) "Weapons of math destruction: How big data increases inequality and threatens democracy" (Penguin Random House LLC, New York).

Parisi, M. L., F. Schiantarelli, A. Sembenelli (2006) "Productivity, innovation and RD: Micro evidence for Italy". European Economic Review, 50(8): 2037–2061.

Phelps, E. S. (1972) "The Statistical Theory of Racism and Sexism". The American Economic Review. 62(4): 659–661.

Porter, C., Serra, D. (2020). Gender Differences in the Choice of Major: The Importance of Female Role Models. American Economic Journal: Applied Economics, 12(3), 226-254.

Profeta, P. (2020) "Gender equality and public policy: Measuring progress in Europe" (Cambridge University Press, Cambridge).

Rhea, A. K., Markey, K., D'Arinzo, L., Schellmann, H., Sloane, M., Squires, P., ... Stoyanovich, J. (2022). "An external stability audit framework to test the validity of personality prediction in AI hiring". Data Mining and Knowledge Discovery, 1-41.

Reskin, B. F. (2001) "Employment discrimination and its remedies". Berg, I., A. L. Kalleberg, eds. Sourcebook of Labor Markets (Springer, Boston), 567–599.

Reskin, B. F., D. Bielby (2005) "A sociological perspective on gender and career outcomes". Journal of Economic Perspectives. 19(1): 71–86.

Ridgeway, C. L. (1997) "Interaction and the conservation of gender inequality: Considering employment". American Sociological Review. 62(2): 218–235.

Ridgeway, C. (2001). Gender, Status, and Leadership. Journal of Social Issues, 57(4), 637-655.

Ridgeway, C. L. (2011) "Framed by Gender: How Gender Inequality Persists in the Modern World" (Oxford University Press, New York).

Ridgeway, C. L. (2019) "Status: Why is it everywhere? Why does it matter?". Russell Sage Foundation (New York).

Ridgeway, C. L., C. Bourg (2004) "Gender as Status: An Expectation States Theory Approach". A. H. Eagly, A. E. Beall, R. J. Sternberg, eds. The psychology of gender (The Guildford Press, New York), 217–241.

Ridgeway, C. L., S. J. Correll (2004) "Motherhood as a status characteristic". Journal of Social issues. 60(4): 683–700.

Robert, L. P., C. Pierce, L. Marquis, S. Kim, R. Alahmad (2020) "Designing fair AI for managing employees in firms: a review, critique, and design agenda". Human-Computer Interaction. 35(5-6): 545-575.

Rosenbaum, P. R. (2002) "Sensitivity to Hidden Bias". P.R. Rosenbaum, eds. Observational Studies (Springer, New York), 71-104.

Rosenbaum, P. R., D. B. Rubin (1985) "Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity Score". The American

Statistician. 39(1): 33-38.

Roth, L. M. (2007) "Women on Wall Street: Despite Diversity Measures, Wall Streets Remains Vulnerable to Sex Discrimination Charges". Academy of Management Perspectives. 21(1): 24–35.

M. Sauder, F. Lynn, J. M., Podolny (2012) "Status: Insights from Organizational Sociology". Annual Review of Sociology 38: 267–283.

Silberg, J., J. Manyika (2019) "Notes from the AI frontier: Tackling bias in AI (and in humans)". McKinsey Global Institute. 1(6).

Stone, P. (2007) "Opting Out? Why Women Really Quit Careers and Head Home" (University of California Press, Berkeley).

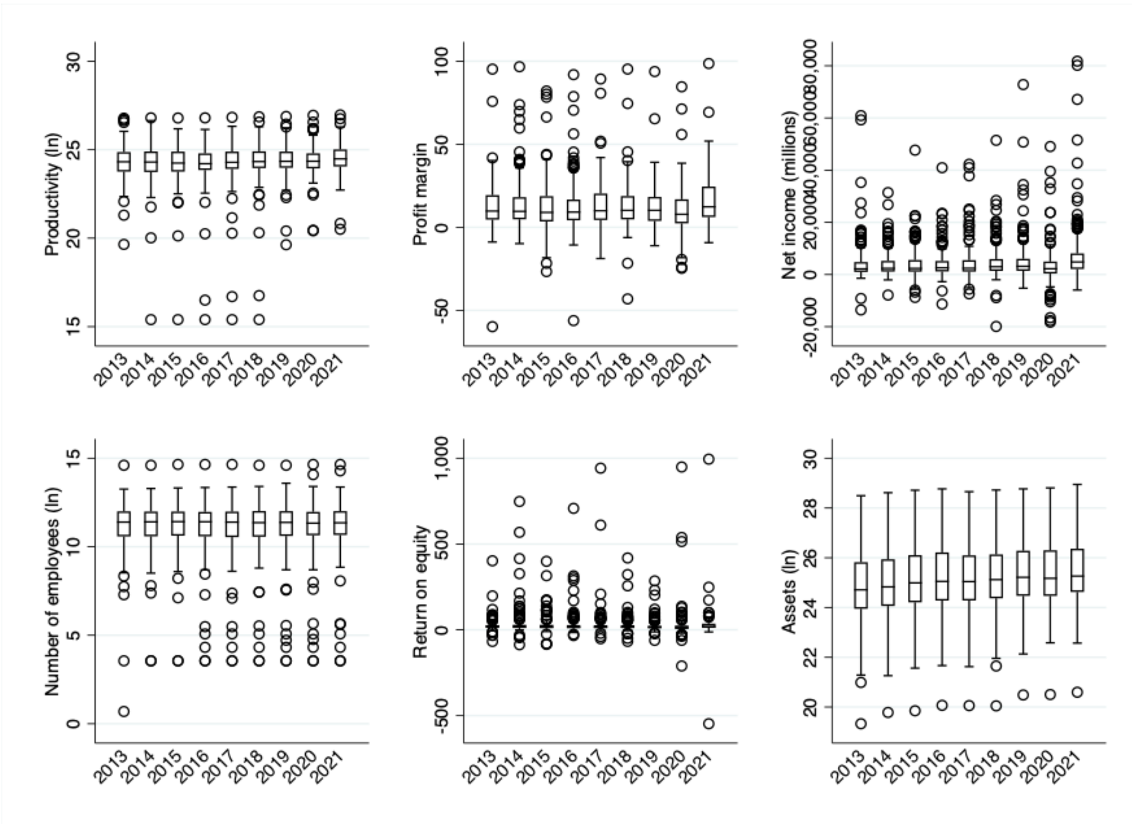
vanEsch, P., S. J. Black, J. Ferolie (2019) "Marketing AI recruitment: The next phase in job application and selection". Computers in Human Behavior. 2019(90): 215–222.

Vella, F. (1994) "Gender Roles and Human Capital Investment: The Relationship between Traditional Attitudes and Female Labor Market Performance". Economica. 61(242): 191–211.

Williams, M. (2022). "Transforming Hiring and Retention: Mr. Edison Meets AI/ML". Harvard Data Science Review. 4(1).

Appendix

Distribution of the observable covariates



Balancing of observable covariates

Before discussing the $ATT(g,t)$ results, estimated through the staggered difference-in-differences approach, this section reports evidence regarding the balance of the observable covariates x among treated and control firms before and after matching. In particular, Figure 6 presents the balancing achieved after the Mahalanobis distance matching.

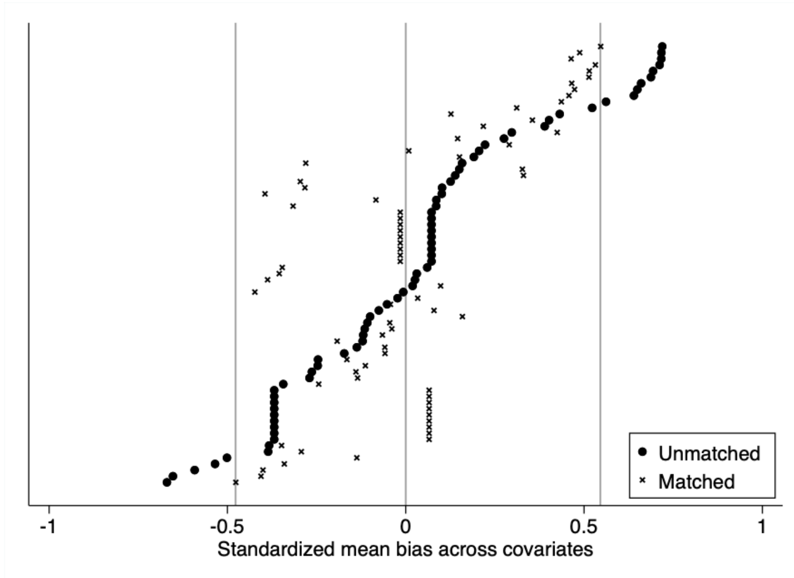


Figure 6: Balancing of the observable covariates achieved after the Mahalanobis distance matching.

Figure 6 shows the standardized mean bias for all covariates before and after matching, that is the difference of the means in the treated and non-treated firms as a percentage of the square root of the average of the sample variances in the treated and non-treated groups (Rosenbaum and Rubin, 1985). The mean absolute standardized bias across covariates after matching is 21.6, which is smaller than the absolute standardized mean bias across covariates before matching (29). As Figure 6 shows, matching reduced the standardized mean bias to less than ∼0.5 for all covariates. Refinement is desirable, but matching has done well at balancing the treated firms and their control counterparts, adjusting reliably for all the covariates.