

Field Experimental Evidence on Gender Discrimination in Hiring: Biased as Heckman and Siegelman Predicted?

Stijn Baert

Abstract

Correspondence studies are nowadays viewed as the most compelling avenue to test for hiring discrimination. However, these studies suffer from one fundamental methodological problem, as formulated by Heckman and Siegelman (*The Urban Institute audit studies: Their methods and findings*. In M. Fix, and R. Struyk (Eds.), *Clear and convincing evidence: Measurement of discrimination in America*, 1993), namely the bias in their results in case of group differences in the variance of unobserved determinants of hiring outcomes. In this study, the authors empirically investigate this bias in the context of gender discrimination. The authors do not find significant evidence for the feared bias.

JEL J16 J71 M51 J41 C93

Keywords Correspondence experiments; Gender discrimination; Unobserved heterogeneity

Authors

Stijn Baert, ✉ Ghent University – Department of Social Economics, Stijn.Baert@UGent.be

Citation Stijn Baert (2015). Field Experimental Evidence on Gender Discrimination in Hiring: Biased as Heckman and Siegelman Predicted? *Economics: The Open-Access, Open-Assessment E-Journal*, 9 (2015-25): 1—11. <http://dx.doi.org/10.5018/economics-ejournal.ja.2015-25>

1 Introduction

During the last decade, economists have attempted to estimate hiring discrimination against women in the labour market by means of correspondence experiments.¹ Within these experiments, pairs of fictitious job applications, only differing by the gender of the candidate, are sent to real job openings. By means of standard probit regressions of the subsequent call-back from the employer on the gender of the candidate, discrimination is identified. The correspondence testing methodology is the golden standard to estimate hiring discrimination in the labour market. It allows to disentangle employer discrimination from supply side determinants of labour market outcomes. Selection on gender differences in (the average level of) unobservable characteristics is not an issue as all the employees' individual characteristics are under control of the researcher (Riach and Rich 2002).

However, a major critique on this methodology can be formulated based on Heckman and Siegelman (1993). They show that not controlling for group differences in the variance of unobservable productivity determinants (and ipso facto of unobservable determinants of positive call-back) can lead to spurious evidence of discrimination. The robustness of ethnic discrimination to the Heckman and Siegelman critique (henceforth "HS critique") is tested by three former contributions to the empirical discrimination literature (Baert et al. 2015, on Belgian data; Carlsson et al. 2014, on Swedish data; Neumark 2012, on US data). These studies show that the HS critique is relevant. The bottom-line of their results is that a higher (perceived) variance in unobservable determinants of positive call-back among ethnic minorities (compared to the ethnic majority) leads to an underestimation of the level of discrimination against them when not controlling for ethnic group differentials in this variance.²

¹ See, e.g., Albert et al. 2011, for Spain; Petit 2007, for France; Riach and Rich 2006, for the UK. Besides its application in studies identifying gender discrimination in hiring, economists have used the correspondence testing framework to test for unequal treatment in the labour market on grounds such as ethnicity, sexual orientation, former unemployment and former employment in the army (see, e.g., Baert 2014; Baert and Balcaen 2013; Baert et al. 2015; Bertrand and Mullainathan 2004; Drydakis 2009; Eriksson and Rooth 2014; Kroft et al. 2013).

² The results presented by Carlsson et al. (2014) deviate to some extent from this empirical pattern.

At the same time, as argued by Azmat and Petrongolo (2014) in their overview of experimental advances in the study of gender differences in the labour market “it should be stressed that existing [...] correspondence evidence on gender discrimination is [...] still open to this criticism.” The only attempt to fill this gap we are aware of, is Carlsson et al. (2014) who apply Neumark’s (2012) econometric framework to a number of already published correspondence studies among which one targeted at gender discrimination. In the present study, we complement their evidence by an empirical investigation of the HS critique in the context of gender discrimination using the same framework but another – and in our opinion theoretically more convincing – identifying assumption.

2 Methods

2.1 Heckman and Siegelman’s Critique

As argued above, correspondence studies adequately address concerns of individual differences in unobservable determinants of productivity. Heckman and Siegelman (1993) show, however, that group differences in the *variance* of these unobservable determinants may still lead to spurious evidence of discrimination.

To see this more clearly for the case of gender discrimination in hiring, assume that both the average observed and the average unobserved determinants of productivity are the same for male and female candidates for an unfilled vacancy, but that the variance of unobservable job-relevant characteristics is, at least in the perception of the employer, higher for females than for males. In addition, suppose that the employer considers the observed determinants of productivity, inferred from the CV and the motivation letter, as relatively low compared to the job requirement. In that case it is rational for the employer to invite the female and not the male candidate, since it is more likely that the sum of observed and unobserved productivity is higher for the female candidates. A correspondence test that detects discrimination against females could therefore underestimate the extent of discrimination against females.³

³ With other assumptions the bias may be in the opposite direction.

2.2 Neumark's Empirical Framework

Neumark (2012) explicitly addresses this critique and provides a statistical procedure to recover unbiased estimates of discrimination. In what follows, we succinctly describe Neumark's approach applied to gender discrimination.

It is well known that in a standard probit model only the ratio of the coefficients to the standard deviation of the unobserved residual is identified. Usually, this standard deviation is arbitrarily set to 1. In our case this means that the variance of unobservable job-relevant characteristics is implicitly assumed to be equal for both males and females, which, for reasons stated above, may bias the measures of discrimination.

Neumark (2012) shows, however, that if the researcher observes job-relevant characteristics that affect the male and female populations' propensities of call-back in the same way, one can identify the ratio of the standard deviation of the unobserved productivity components of these groups.⁴ Implementing Neumark's (2012) idea in the context of gender discrimination boils down to the estimation of a heteroskedastic probit model in which the variance of the error term is allowed to vary with gender.

2.3 Identification Strategy

As mentioned in the previous subsection, identification of the group-specific variance in observable determinants of positive call-back within the heteroskedastic probit framework requires experimental data with variation in observable job-relevant characteristics that affect the (in our case gender) groups' propensities of call-back in the same way. Variables used by Baert et al. (2015), Carlsson et al. (2014) and Neumark (2012) in their application of the Neumark framework in the context of ethnic discrimination were education level, personality traits, work experience, type of neighbourhood, sport activities and application quality. In the context of gender discrimination, Carlsson et al. (2014) assumed equal returns for both genders from variation in educational degree,

⁴ The intuition is that if in a standard probit model the estimated coefficients of these job-relevant characteristics differ by gender, then this must be a consequence of a differential standard deviation, since by assumption the coefficient of these characteristics should be the same across groups (and since, as mentioned before, in a probit model only the ratio of the coefficients to the standard deviation are identified).

international mobility, work experience, employment status and job tenure. Their choice can be criticised on theoretical grounds.⁵ All the aforementioned variables used for identification of the Neumark procedure result from variation in choices and outcomes at the employee side. Therefore, they may be correlated with ethnicity or gender in reality.⁶

The alternative variable we assume to have the same return across groups is the distance between the candidate's living place and the workplace. On the one hand, it is clear that this variable has the potential to affect hiring decisions of employers. This is the case as employers may prefer workers with a social network in the neighbourhood of the firm. In addition, they may expect a higher commitment from workers living close to the firm (and, therefore, wasting not too much time by commuting). On the other hand, by using this variable we actually exploit employer variation instead of employee variation as the living place of the employee is constant. As a result, there is no reason why this variable would be more rewarded for members of a particular sex.⁷ Both considerations are confirmed empirically (see Section 4).

3 Data

We use data from Baert et al. (Forthcoming), a correspondence study investigating the importance of employer preferences in explaining Sticky Floors. Sticky Floors are defined as the pattern that women are, compared to men, less likely to start to

⁵ It should be noted, however, that based on the empirical tests the aforementioned authors present one cannot reject that the chosen variables affect call-back probabilities with a different magnitude for the groups they study.

⁶ For instance, members of ethnic minorities may have a higher probability of living in more disadvantaged neighbourhoods (Bertrand and Mullainathan 2004). As particular values for the aforementioned variables may (not) square (and therefore enforce or disprove) prejudices about ethnic minorities or women, variation in these variables may be expected to be valued differently for these groups.

⁷ One could argue that applications to employers living very far away from the residence of the applicant reflect a willingness to be mobile which may be correlated with female sex. Women of child-bearing and rearing age might be perceived as being less flexible when it comes to the distance between their workplace and residence. However, the fictitious job candidates in the experimental data we use mentioned they were quite young (26 or 27), unmarried and not having children (see Section 4). Therefore, we assume that the return to living close to the workplace is the same for the candidates of both genders within our data. Moreover, if we redo our estimations using only observations with distances lower than 30 minutes of car driving, the results are very comparable to the ones presented in the main text.

climb the job ladder. To this end, these authors sent fictitious job applications to real job openings in the labour market of Flanders between October 2013 and March 2014. During this period, they randomly selected 288 vacancies for jobs targeting Bachelors in business administration and 288 vacancies for jobs targeting Masters in business economics in the private sector. They restricted themselves to vacancies requiring at most five years of work experience. Two job applications of individuals with five years of work experience (in a first and current job), identical in terms of productivity-relevant characteristics, were sent to the selected vacancies. These applicants were single individuals born, studying and living in comparable suburbs of Ghent, the second largest city of Flanders. Within each pair of applicants, a typically male sounding name was randomly assigned to one of both applications and a typically female sounding name to the other one. Call-backs were received via telephone voicemail and email.

Baert et al. (Forthcoming) sent applications both to vacancies implying a promotion in terms of occupational level and/or job authority and to vacancies at the same level. Thereby, they were able to test whether unequal treatment of young men and women in hiring was heterogeneous by whether or not jobs implied a promotion in comparison with employees' current position. They found significant evidence of hiring discrimination against females when they applied for jobs at a higher occupational level. For these jobs, females got, compared to males, about 33% less invitations for a job interview and 19% less positive reactions in broad sense. On the other hand, they found no significant heterogeneity in hiring discrimination by the job authority level of the posted jobs.

In the present study, we will test whether the discrimination measures presented by Baert et al. (Forthcoming) are biased by gender differences in the variance of unobserved determinants of hiring outcomes. Therefore, the data from Baert et al. (Forthcoming) are, in view of our mentioned identifying strategy, extended with the distance between the workplace announced in the vacancy and the candidate's residence.⁸

⁸ This distance, expressed in minutes when driving by car, is calculated using the online routing tool of Google Maps.

4 Results

Table 1 presents the results of our empirical analysis. In Panel A we report the degree of gender discrimination that comes out of a standard analysis of the data of Baert et al. (Forthcoming). We retake their main findings by conducting basic probit estimations with positive call-back as an outcome variable. Positive call-back is defined as getting an invitation for an interview concerning the announced job in models (1) and (2) and defined as getting any positive reaction from the employer side in models (3) and (4).

On the one hand (in models (1) and (3)), we regress positive call-back on a dummy indicating female sex of the candidate and the distance between the workplace and the residence of the applicant. On the other hand, for models (2) and (4), the effect of female sex is broken down by whether the vacancy indicated a job implying a promotion in occupational level compared with the current job of the candidate. This is done by replacing the dummy indicating female sex of the candidate by two dummies: one indicating female candidates who applied for a job not implying a promotion in occupational level and one indicating female candidates who applied for a job implying a promotion in occupational level.⁹

By doing that, we get results that are very similar to those presented in Table 4 and Table 5 of Baert et al. (Forthcoming). More concretely, the regression results indicate that, overall, the tested employers did not discriminate based on sex. However, if the effect of revealing female gender is broken down by the occupational level of the posted job, we find that a female name lowers the probability of positive call-back by four to five percentage points when one applies for jobs implying a promotion in this respect.

Interestingly, the estimation results for the variable “distance between the workplace and the candidate’s residence” – not presented in Table 1 but available on request – are, for all of the mentioned models, highly significantly different from zero ($p < 0.01$) and have the expected (negative) sign. Moreover, based on a Wald test applied to the estimation results of an alternative probit model with an additional interaction variable between female sex of the candidate and the distance between workplace and residence, we cannot reject that this distance

⁹ This operation also implies the introduction of a dummy indicating promotion jobs in terms of occupational level and a dummy indicating promotion jobs in terms of job authority without an interaction with female sex.

variable is rewarded equally for males and females. The test results are summarised in Table 1.

Panel B reports the results based on a re-estimation of models (1) to (4) by means of a heteroskedastic probit model in the spirit of Neumark (2012) allowing the variance of the error term to vary with the gender of candidate. By doing that, we get unbiased results that are very comparable to those in Panel A. In other words: we find no evidence for a bias in the sense of the HS critique. This finding is related to the fact that the estimated male and female standard deviations concerning the error term (σ_{Male} and σ_{Female}) are very comparable. Therefore, our results seem to indicate that the tested employers do not perceive a (gender) group difference in the variance of unobserved determinants of productivity. These results, therefore, corroborate with those of Carlsson et al. (2014) based on correspondence testing data gathered in Sweden.

Last, we decompose, in the spirit of Neumark (2012), the unbiased estimates in an effect through level (keeping group differences in the variance of the error term constant) and an effect through variance (keeping differences in unbiased parameters constant). Interestingly, but differing from the findings of Carlsson et al. (2014), we find that the effects through level are, although not significantly different from zero, more or less of the same magnitude as the total unbiased effect. In addition, the effects through variance are rather close to zero.

Our result of no important perceived gender group difference in the variance of unobserved variables deviates from the finding of the more substantial ethnic group difference in this respect outlined in Baert et al. (2015), Carlsson et al. (2014) and Neumark (2012). One explanation for this finding is that perceived group differences in the variance of unobserved variables can be thought of as a sort of statistical discrimination. Following Altonji and Blank (1999) employers may believe that the same observable signal is more precise for one group compared to another. This theory seems to be more applicable to ethnic groups than to gender groups.

5 Conclusion

In this study, we investigated the research gap indicated by Azmat and Petrongolo (2014). This gap boils down to the fact that standard analyses of correspondence

testing data aimed at investigating hiring discrimination do not control for group differences in the variance of unobservable productivity determinants and, as a consequence of that, may be biased. While the robustness of ethnic discrimination to the this critique, formulated first by Heckman and Siegelman (1993), is tested by three former studies, Azmat and Petrongolo (2014) stress that correspondence studies on gender discrimination are still open to this critique. Estimating the bias feared by Heckman and Siegelman (1993) in the context of gender discrimination was the aim (and the contribution) of this study.

We used Belgian correspondence data aimed at measuring hiring discrimination against young females. We employed the empirical framework introduced by Neumark (2012) and proposed an original identifying assumption. By doing that, we found no significant evidence for the bias feared by Heckman and Siegelman (1993) related to the fact that the estimated (perceived) variance of unobservables is very comparable for male and female job candidates.

The issue of gender differences in heterogeneity with respect to productivity is an important puzzle piece in the study of gender convergence in the labour market. We contribute modestly to this literature by showing that, at least in the perception of Belgian employers, there is no evidence for the hypothesis that women are (perceived as) more heterogeneous than men in productivity related variables unobservable to researchers.

References

- Albert, R., L. Escot, and J. Fernández-Cornejo (2011). A field experiment to study sex and age discrimination in the Madrid labour market. *International Journal of Human Resource Management* 22 (2): 351–375.
URL: <http://dx.doi.org/10.1080/09585192.2011.540160>.
- Altonji, J., and R. Blank (1999). Race and Gender in the Labor Market. In O. Ashenfelter, and D. Card (Eds.), *Handbook of Labor Economics*. Amsterdam: Elsevier.
- Azmat, G., and B. Petrongolo (2014). Gender and the Labor Market: What Have We Learned from Field and Lab Experiments? *Labour Economics* 30 (Special Issue on “What determined the dynamics of labour economics research in the past 25 years?”): 32–40. URL: <http://dx.doi.org/10.1016/j.labeco.2014.06.005>.
- Baert, S. (2014). Career lesbians. Getting hired for not having kids. *Industrial Relations* 45 (6): 543–561. URL: <http://dx.doi.org/10.1111/irj.12078>.

- Baert, S., and P. Balcaen (2013). The Impact of Military Work Experience on Later Hiring Chances in the Civilian Labour Market. Evidence from a Field Experiment. *Economics: The Open-Access, Open-Assessment E-Journal* 7 (2013-37): 1–17. URL: <http://dx.doi.org/10.5018/economics-ejournal.ja.2013-37>.
- Baert, S., B. Cockx, N. Gheyle, and C. Vandamme (2015). Is There Less Discrimination in Occupations Where Recruitment Is Difficult? *Industrial and Labor Relations Review* 68 (3): 467–500. URL: <http://ilr.sagepub.com/content/68/3/467>.
- Baert, S., A. De Pauw, and N. Deschacht (Forthcoming). Do Employer Preferences Contribute to Sticky Floors? *Industrial and Labor Relations Review*. URL: <https://ideas.repec.org/p/iza/izadps/dp8447.html>.
- Bertrand, M., and S. Mullainathan (2004). Are Emily and Greg More Employable than Lakisha and Jamal? A field experiment on labor market discrimination. *American Economic Review* 94 (4): 991–1013. URL: <http://dx.doi.org/10.1257/0002828042002561>.
- Carlsson, M., L. Fumarco, and D.-O. Rooth (2014). Does the design of correspondence studies influence the measurement of discrimination? *IZA Journal of Migration* 2014 (3): 11. URL: <http://dx.doi.org/10.1186/2193-9039-3-11>.
- Drydakis, N. (2009). Sexual orientation discrimination in the labour market. *Labour Economics* 16 (4): 364–372. URL: <http://dx.doi.org/10.1016/j.labeco.2008.12.003>.
- Eriksson, S., and D.-O. Rooth (2014). Do Employers Use Unemployment as a Sorting Criterion When Hiring? Evidence from a Field Experiment. *American Economic Review* 104 (3): 1014–1039. URL: <http://dx.doi.org/10.1257/aer.104.3.1014>.
- Heckman, J.J., and P. Siegelman (1993). The Urban Institute audit studies: Their methods and findings. In M. Fix, and R. Struyk (Eds.), *Clear and convincing evidence: Measurement of discrimination in America*. Washington DC: Urban Institute Press.
- Kroft, K., F. Lange, and M.J. Notowidigdo (2013). Duration Dependence and Labor Market Conditions: Evidence from a Field Experiment. *Quarterly Journal of Economics* 128 (3): 1123–1167. URL: <http://dx.doi.org/10.1093/qje/qjt015>.
- Neumark, D. (2012). Detecting discrimination with audit and correspondence studies. *Journal of Human Resources* 47 (4):1128–1157. URL: <http://jhr.uwpress.org/content/47/4/1128.abstract>.
- Petit, P. (2007). The effects of age and family constraints on gender hiring discrimination: A field experiment in the French financial sector. *Labour Economics* 14 (3): 371–391. URL: <http://dx.doi.org/10.1016/j.labeco.2006.01.006>.
- Riach, P.A., and J. Rich (2002). Field experiments of discrimination in the market place. *Economic Journal* 112 (November): F480–F518. URL: <http://dx.doi.org/10.1111/1468-0297.00080>.

Riach, P.A., and J. Rich (2006). An Experimental Investigation of Sexual Discrimination in Hiring in the English Labor Market. *B.E. Journal of Economic Analysis & Policy (Advances)* 5 (2): 1. URL: <http://dx.doi.org/10.2202/1538-0637.1416>.

Table 1: Estimation results

	Model (1)	Model (2)	Model (3)	Model (4)
A. Estimates from basic probit model				
Female candidate	-0.010 (0.013)		-0.010 (0.017)	
Female candidate x No promotion in occupational level		0.022 (0.018)		0.042 (0.026)
Female candidate x Promotion in occupational level		-0.040** (0.019)		-0.050** (0.022)
B. Estimates from heteroskedastic probit model				
Female candidate	-0.012 (0.020)		-0.011 (0.017)	
Female candidate x No promotion in occupational level		0.022 (0.021)		0.041 (0.029)
Female candidate x Promotion in occupational level		-0.037** (0.017)		-0.050** (0.020)
C. Effect through level				
Female candidate	-0.003 (0.049)		0.010 (0.051)	
Female candidate x No promotion in occupational level		0.033 (0.053)		0.063 (0.059)
Female candidate x Promotion in occupational level		-0.027 (0.046)		-0.028 (0.049)
D. Effect through variance				
Female candidate	-0.008 (0.042)		-0.020 (0.042)	
Female candidate x No promotion in occupational level		-0.011 (0.046)		-0.022 (0.036)
Female candidate x Promotion in occupational level		-0.010 (0.041)		-0.023 (0.045)
Log ($\sigma_{\text{Female}}/\sigma_{\text{Male}}$)	-0.052	-0.041	-0.110	-0.105
Wald test statistic, null hypothesis that $\sigma_{\text{Female}}/\sigma_{\text{Male}} = 1$ (p-value)	0.852	0.811	0.612	0.596
Wald test statistic, null hypothesis that ratio of coefficients for distance between workplace and residence = 1 (p-value)	0.910	0.885	0.732	0.721
Dependent variable: invitation to a job interview	Yes	Yes	No	No
Dependent variable: any positive reaction	No	No	Yes	Yes
Observations	1152	1152	1152	1152

Notes: Additional controls included in the basic probit and heteroskedastic probit models are: distance between the workplace and the candidate's residence and, for models (2) and (4) a dummy that indicates jobs implying a promotion. The presented statistics are marginal effects and standard errors, corrected for clustering at the vacancy level, in parentheses. *** (***) ((**)) indicates significance at the 1% (5%) ((10%)) level.

Please note:

You are most sincerely encouraged to participate in the open assessment of this article. You can do so by either recommending the article or by posting your comments.

Please go to:

<http://www.economics-ejournal.org/economics/journalarticles/2015-25>

The Editor