

Manuscript ECONOMICS-KIEL-2015-44 | Response to Referee Report 2

I sincerely thank the referee for her/his constructive comments, helping me to improve the quality of my manuscript. Below one can find responses to the points raised in the referee report. The original comments of the referee are in italics and shaded in grey.

Comments on the manuscript MS 1516 entitled 'Field experimental evidence on gender discrimination in hiring: Biased as Heckman and Siegelman predicted?'

I have few comments as I think the paper is well written and the analysis is competent.

I don't think 'predict' is quite right for Heckman and Siegelman's comments though. These comments were more a critique of the audit approach in field experiment testing – that is using actual testers to go in person to make job enquiries or undertake job interviews. As such they argued the researcher could not control for all aspects of the applicant's characteristics that could influence an employer. I don't feel Heckman and Siegelman were predicting the responses would be biased, more the experiment could overestimate discrimination if a tester influenced the employer.

Heckman and Siegelman's critique boils down to the fact that field experiments on discrimination may fail to ensure that fictitious job candidates from different groups (in my application: applicants of a different gender) appear identical to employers. As mentioned by the referee, this problem is more important for audit than for correspondence studies. While in the former studies even *average* observed and unobserved productivity characteristics may differ between groups, this is not the case in correspondence studies. However, as argued in my manuscript, and as proven rigorously by the cited work of Neumark (2012), Heckman and Siegelman's critique applies even in the case of correspondence studies due to differences (as perceived by employers) in the *variance* of unobservables between, for instance, men and women. This might bias the measure of discrimination based on correspondence experiments. To see this more clearly, I refer to the following paragraph of my initial manuscript (page 4, from line 9 on):

“[...] assume that both the average observed and 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.”

Having said that, I acknowledge that the word “predicted” in the title of the initial manuscript might be an unfortunate choice. If the editor agrees with this, I would change this word into “feared” (both in the title and throughout the article).

The main comment I have is that details as to the actual experiment are needed given that the focus of the paper is on explaining whether any bias resulted from the experiment that was conducted. When was the study conducted? Where? What were the actual response numbers?

As mentioned in my response to Referee Report 1, in the case of acceptance, I will prepare a revised version of the manuscript in which a paragraph is added with more information about the data gathering and randomisation process. Concerning the referee’s punctual questions: the experiment was conducted (i) between October 2013 and March 2014, (ii) in Flanders, the Northern part of Belgium and (iii) male candidates got in 149 of their 576 applications a positive reaction while female candidates got in 144 cases a positive reaction.

As to the main finding – no significant evidence for the predicted bias – it may be better expressed as ‘no under or over estimation of discrimination was evident’.

In line with my response to the first comment of the referee, I propose to change the word “predicted” into the word “feared” in the mentioned sentence. I believe this (revised) sentence fits the paragraph’s flow better than the one proposed by the referee.

Expression again in the conclusion - - rather than ‘occupied’ say ‘investigated’; rather than ‘gap’ write ‘critique’.

In line with my response to the first and the third comment of the referee, I propose to change the word “predicted” into the word “feared” in the mentioned sentence. In addition, I agree to change the word “occupied” into “investigated”. Concerning the word “gap”, however, I propose to keep this word as it refers to the research gap mentioned by Azmat and Petrongolo (2014) – these authors do not formulate a critique.

References

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.

Neumark, D. (2012). Detecting discrimination with audit and correspondence studies. *Journal of Human Resources* 47 (4):1128–1157.