

Extended Abstract: “Gender Bias in Dressage Judging”

Anna Sandberg, Stockholm School of Economics, anna.sandberg@hhs.se

In most countries, gender equality on the labor market has increased substantially during the past decades. However, the gender wage gap persists and the presence of females in high-ranking positions is still negligible. Some argue that in order to break this “glass ceiling”, gender quotas for top positions need to be introduced, because once women reach top positions they will hire more women than their male counterparts. However, there is little empirical evidence on whether women are more likely to be hired or positively evaluated by women than by men, and the few existing studies show ambiguous results (see e.g. Bagues and Esteve-Volart, 2010; Zinovyeva and Bagues, 2011; De Paola and Scoppa, 2011; Broder, 1993; Abrevaya and Hamermesh, 2012). More research is clearly needed to deepen our understanding of when, why and how the gender of the evaluator matters.

Empirical evidence on this topic is hard to come by since in most recruitment or promotion procedures the gender composition of evaluation committees are systematically related to the gender composition and quality of the candidates being evaluated. Also, even in the case of random assignment of evaluators to evaluating committees, researchers can rarely observe the individual decisions made by each member of an evaluation committee. To empirically investigate interaction effects between the gender of the candidate and the gender of the evaluator, ideally one needs a large dataset of comparable decisions where each candidate is evaluated by evaluators of both genders, and the opinion of each individual evaluator can be quantified. While such data are almost impossible to obtain from a regular labor market setting, they can be obtained from the equestrian sport dressage.

During the past few years, Economists have increasingly turned to sports data to investigate for instance racial discrimination, nationalistic bias, home-team favoritism and order effects (e.g. Garicano et al., 2005; Zitzewitz 2006; Dohmen, 2008; Pettersson-Lidbom and Priks, 2010). Recent research on racial discrimination in sports points to an interaction effect between the ethnicity of the judge and the ethnicity of the athlete. Price and Wolfers (2010) find that NBA players receive fewer fouls when more referees of their own race are included in the referee crew. When investigating the behavior of umpires in Major League Baseball, Parsons, Sulaeman, Yates and Hamermesh (2011) find similar results. However, no previous study uses sports data to investigate gender bias in judging. I fill this gap by taking advantage of the unique features of dressage competitions.

Dressage is the only major judging sport where male and female athletes compete directly against each other on equal terms and are judged by both male and female judges. During dressage competitions, each rider performs a series of dressage movements, and is judged based on the technical quality of the movements and the overall artistic quality of the performance. The judging panel consists of five judges, and the final position of each rider is determined by the average of the scores given by each of the five judges.

In this study, I investigate gender effects in dressage competitions using a unique dataset containing detailed judging results from international dressage competitions. For each

performance given by a rider, I have access to the technical and artistic scores given by each judge, allowing me to compare the scores given by male and female judges who observe the exact same performance. In total, I have access to 96,087 scores given by international dressage judges between the years 2007 and 2012. The dataset contains scores given to 1224 individual riders (837 women, 387 men) by 192 individual judges (106 women, 86 men). The fairly equal gender distribution of riders and judges in professional dressage ensures a relatively large sample size of performances evaluated by both male and female judges.

First, I test for an interaction effect between the gender of the judge and the gender of the rider, i.e. I investigate if riders are evaluated more or less harshly depending on whether the judge is of the same or of the opposite gender as the rider. To do this I regress the total score given by a judge on the gender of the judge, including performance fixed effects and running separate regressions for female and male riders. I find no evidence of an interaction effect between the gender of the judge and the gender of the rider. However, when performing the same analysis for nationality instead of gender, I find a large and significantly positive interaction effect between the nationality of the judge and the nationality of the rider, indicating a strong nationalistic bias. Thus, in this type of setting (mixed-gender competitions with non-anonymous face-to-face judging procedures, where evaluators are formally trained in judging according to objective standards and face large career incentives to be accurate in their judging) nationalistic bias seems to be more persistent than gender bias.

During the spring I will continue analyzing this data set, looking at potential heterogeneous effects of judge and rider gender. In addition to judge and rider gender, I have access to a rich set of background variables such as the age and nationality of judges and riders and the monthly world ranking of each rider. I also plan to collect data on judge experience and rider attractiveness, and investigate the impact of these variables on gender differences in judging behavior.

I will also investigate peer effects among the group of judges. Most international dressage competitions consist of three different performances, and all five judges score each performance. When all riders have completed their first performance, the judges can talk to each other and observe each other's scores. I will look into whether, when scoring the second performance of a rider, the judges are influenced by how the other judges scored the first performance of that rider. In particular, I am interested in investigating gender differences in whether and how judges conform to the opinion of the other judges on the panel.

References:

Abrevaya, J. and Hamermesh, D., (2012), "Charity and Favoritism in the Field: Are Female Economists Nicer (to Each Other)?", *Review of Economics and Statistics*, 94, 202-7.

Bagues, M. and Esteve-Volart, B., (2010), "Can Gender Parity Break the Glass Ceiling? Evidence from a Repeated Randomized Experiment", *Review of Economic Studies*, 77, 1301-1328.

Broder, I.E., (1993), "Review of NSF Economics Proposals: Gender and Institutional Patterns", *American Economic Review*, 83(4), 964-970.

Dohmen, T.J., (2008), “The Influence of Social Forces: Evidence from the Behavior of Football Referees”, *Economic Inquiry*, 46(3), 411–24.

Garicano, L., Palacios-Huerta, I. and Prendergast, C., (2005), “Favoritism under Social Pressure”, *Review of Economics and Statistics*, 87(2), 208–16.

De Paola, M. and Scoppa, V., (2011), ”Gender Discrimination and Evaluators’ Gender: Evidence from the Italian Academy”, Working Paper n. 06-2011, Dipartimento di Economia e Statistica, Università della Calabria.

Parsons, Christopher, Johan Sulaeman, Michael Yates, and Daniel Hamermesh, (2011), “Strike Three: Discrimination, Incentives and Evaluation,” *American Economic Review*, 101, 1410–1435.

Pettersson-Lidbom, P. and Priks, M., (2010), “Behavior under social pressure: Empty Italian stadiums and referee bias”, *Economics Letters*, 108(2), 212-214.

Price, Joseph, and Justin Wolfers, “Racial Discrimination among NBA Referees,” (2010), *Quarterly Journal of Economics*, 125, 1859– 1887.

Zinovyeva, N. and Bagues, M., (2011), “Does gender matter for academic promotion? Evidence from a randomized natural experiment”, IZA Discussion Paper 5537.

Zitzewitz, E., (2006), “Nationalism in Winter Sports Judging and It’s Lessons for Organizational Decision Making”, *Journal of Economics and Management Strategy*, 15(1), 67–99.