

The Gender Pay Gap Revisited: Methodological Improvements with Big Data*

- preliminary -

Anthony Strittmatter[†] Conny Wunsch[‡]

November 26, 2019

Abstract

The vast majority of existing studies that estimate the gender pay gap uses one of the versions of the Blinder-Oaxaca decomposition. We use a very large data set of 1.7 million employees in Switzerland and study how methodological improvements that are possible with such big data affect estimates of the gender pay gap. First, we study the sensitivity of estimates with respect to the availability of observationally comparable men and women. Second, we compare the estimates obtained from different parametric, semi-parametric and non-parametric estimators for the wage gap, including variants that make use of machine learning methods. Finally, we investigate whether heterogeneity in wage gaps is a possible explanation for different results. We find that common support breaks down quickly and that enforcing support changes the results obtained from the parametric estimators. Moreover, the semi-parametric estimators, in particular combinations of exact and propensity score matching, yield considerably lower unexplained wage gaps than the Blinder-Oaxaca technique. We also find that average pay gaps hide important heterogeneity and that estimated heterogeneity itself depends on the method. We conclude that the choice of method matters a lot, suggesting that policy makers should be more careful with basing their decisions on Blinder-Oaxaca estimates.

Keywords: Inequality, Gender Pay Gap.

JEL classification: J31

*We are grateful for helpful comments from seminar participants at LISER, Luxembourg. The authors are solely responsible for the analysis and the interpretation thereof.

[†]Department of Economics, University of St. Gallen; email: anthony.strittmatter@unisg.ch.

[‡]Faculty of Business and Economics, University of Basel; email: conny.wunsch@unibas.ch.