

# Estimating a counterfactual wage heavy-tailed distribution using survey data

Mihaela-Catalina Anastasiade-Guinand\*, Alina Matei†and Yves Tillé‡

Conference in honour of Fred Smith and Chris Skinner

Online, 7–9 July 2021

## Abstract

We work in the framework of the gender wage modelisation using survey data. The wage of an employee is hypothetically a reflection of their characteristics, such as the education level or the work experience. It is possible that a man and a woman with the same characteristics get different salaries. To measure the difference in the gender wages we use the concept of counterfactual distribution. This is done in order to estimate what the former group would earn, if they had the characteristics of the latter group. The usual regression approach of Blinder-Oaxaca consists of modeling the mean of the wage of each individual conditionally on their characteristics. The aim is to isolate the part attributable to gender at the mean level by estimating the part of the wage difference that is explained by the differing characteristics. Conditional to some characteristics, we assume that the conditional wage distribution of each woman follows a given theoretical distribution with unknown parameters. First, we estimate the parameters of the distribution of each woman given their characteristics. Next, the marginal women wages distribution is fitted based on the individual woman wage distributions. A counterfactual distribution is constructed by reweighting the women characteristics. We provide two parametric methods to estimate the gender wage quantiles and counterfactual wage quantiles, respectively, and estimate their differences. The goal is to capture the shape of the wage distributions and to go beyond the simple mean differences, by determining the estimator of ‘gender wage discrimination’ at different quantiles. Since, in general, wage distributions are heavy-tailed, the main

---

\*Swiss Federal Statistical Office

†Institute of Statistics, University of Neuchâtel, Switzerland

‡Institute of Statistics, University of Neuchâtel, Switzerland

interest is to model wages by using heavy-tailed distributions like the GB2 distribution. We illustrate the two proposed methods using the GB2 distribution and real data from the Swiss Federal Statistical Office.