

ESTIMATION D'UNE DISTRIBUTION CONTREFACTUELLE DES SALAIRES À L'AIDE DE DONNÉES D'ENQUÊTE

Alina Matei¹ & Mihaela-Catalina Anastasiade² & Yves Tillé³

¹ *Institut de Statistique, Université de Neuchâtel, Suisse, Av. Bellevaux 51, 2000, Neuchâtel et alina.matei@unine.ch*

² *Office Fédéral de la Statistique, Neuchâtel, Suisse, Espace de l'Europe 10, 2010, Neuchâtel et mihaela-catalina.anastasiade@bfs.admin.ch*

³ *Institut de Statistique, Université de Neuchâtel, Suisse, Av. Bellevaux 51, 2000, Neuchâtel et yves.tille@unine.ch*

Résumé. Le salaire d'un employé est hypothétiquement un reflet de ses caractéristiques, telles que le niveau d'éducation ou l'expérience de travail. Il est possible qu'un homme et une femme ayant les mêmes caractéristiques perçoivent des salaires différents. En nous basant sur des données d'enquête, nous modélisons les distributions de salaires. Pour mesurer la différence entre les salaires d'un homme et d'une femme, nous utilisons une distribution contrefactuelle (voir Fortin et al., 2011). Celle-ci est utile pour estimer ce que gagneraient les femmes, si elles avaient les caractéristiques des hommes. L'approche de régression de Blinder (1973) et de Oaxaca (1973) consiste à modéliser la moyenne du salaire de chaque individu conditionnellement à ses caractéristiques. L'objectif est d'isoler la part attribuable à la discrimination salariale entre hommes et femmes au niveau moyen.

Conditionnellement à certaines caractéristiques, nous supposons que la distribution conditionnelle du salaire de chaque femme suit une distribution théorique donnée avec des paramètres inconnus. Nous estimons les paramètres de la distribution du salaire de chaque femme en fonction de ses caractéristiques. Ensuite, la distribution marginale des salaires des femmes est estimée en fonction des distributions individuelles des salaires des femmes. Une distribution contrefactuelle est construite en pondérant la distribution des salaires des femmes (Anastasiade and Tillé, 2017). Nous proposons deux méthodes paramétriques pour estimer les quantiles des salaires des femmes, des hommes et les quantiles des salaires contrefactuels. L'objectif est de calculer les différences entre ceux-ci pour estimer la discrimination salariale à différents quantiles. Nous illustrons les deux méthodes proposées en utilisant la distribution GB2 utile pour modéliser des salaires et nous les comparons avec d'autres approches données dans la littérature.

Mots-clés. Analyse de données d'enquêtes, distribution contrefactuelle, régression GB2, estimation des quantiles.

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. A counterfactual distribution is estimated by putting together the parameters of one group and the characteristics of the other group (for an overview, see Fortin et al., 2011). This is done in order to estimate what the former group would earn, if they had the characteristics of the latter group. The regression approach of Blinder (1973) and Oaxaca (1973) consists of modelling the mean of the wage of each individual conditionally on their characteristics. The aim is to isolate the part attributable to gender discrimination 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 wage distribution (Anastasiade and Tillé, 2017). 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 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 compare them with other approaches found in the topic-related literature.

Keywords. Analysis of survey data, counterfactual distribution, GB2 regression, quantile estimation.

Bibliographie

- Anastasiade, M.-C. and Tillé, Y. (2017). Decomposition of gender wage inequalities through calibration: Application to the Swiss structure of earnings survey. *Survey Methodology*, 43(2), 211–234.
- Blinder, A. S. (1973). Wage discrimination: Reduced form and structural estimates. *Journal of Human Resources*, 8(4), 436–455.
- Fortin, N., Lemieux, T., and Firpo, S. (2011). Decomposition Methods in Economics, *Handbook of Labor Economics*, volume 4, chapter 1, 1–102. Elsevier.
- Oaxaca, R. (1973). Male-Female Wage Differentials in Urban Labor Markets. *International Economic Review*, 14(3), 693–709.