The Gender Pay Gap Revisited with Big Data: Do Methodological Choices Matter?

Anonymous Author(s)
email

1 Summary

The unexplained gender pay gap is a highly relevant topic for policy makers and analyzed in various papers. In contrast to other studies that investigate the sources of inequality in pay, Strittmatter et al. [4] analyze the impact of different methodological choices on the resulting estimated unexplained pay gap. They investigate three types of methodological choices:

1. **Enforcement of common support**: They define the unexplained gender pay gap as "the expected relative wage difference of employed women in the sample with support, compared to employed men with the same observed wage determinants". However, when doing so, one needs to define what "with support" actually means and there are multiple methods to do so. One can impose support for many variables (resulting in higher accuracy, but a smaller dataset) or only a few. Furthermore, hybrid methods (exact matching for important variables, approximate matching for less important ones) can be used.
2. **Inclusion of variables**: As with other machine learning problems, one needs to decide which variables, interactions, and transformations to include in the model. They analyze a baseline model with only lower-order terms, a full model with much more non-linear and interaction terms, and automatic variable selection with LASSO.
3. **Estimators**: Different estimators for the pay gap can be used. In the paper, they analyze simple linear regression, Blinder-Oaxaca decomposition (allowing different impact of individual variables per gender), inverse probability weighting estimators (creating a pseudo-population that allows to infer causal relationships by weighting the samples accordingly), and matching estimators (dividing the population into mutually exclusive groups according to the similarity of their features).

They use the Swiss Earnings Structure Survey as the basis for their analysis. This large dataset with informations on more than 1.7 million employees allows to use more flexible methods for the estimators and the inclusion of more variables without overfitting.

According to their analysis, the three investigated methodological choices matter a lot and can reduce the unexplained pay gap by up to 50% compared to the commonly used Blinder-Oaxaca estimates with linear models.

2 Critique

Overall, the paper gives interesting insights and quantitatively demonstrates that the methodological choices matter a lot for the estimation of the unexplained gender pay gap. Their exact numbers for the unexplained gender wage gap are also questionable (because of the dataset and because they are very different depending on which model is used), but their goal was not to get a "correct" estimate, but rather show the impact of the methodology, and they succeeded in doing so in my opinion. The figures they provide supplement the text well. I especially liked the analysis of common support by

Submitted as response essay for the *Sequencing Legal DNA: NLP for Law and Political Economy* course at ETH Zurich. Do not distribute.

sector, as this also visualizes according to which factors men and women are most different in the workforce and how this changes when looking at the private or public sector. They also reinforce the most important results with their tables in the appendix, although this was quite hard to find and could be published more prominently.

I will now address different aspects of the paper in more detail.

2.1 Dataset

Although the dataset is relatively large and has many predictors, I think that some important information is still missing. As in other studies of the gender pay gap, one of the most important wage determinants, namely the actual work experience, is missing and relatively poor proxies (that differ a lot by gender) such as age, education, and tenure are used instead. However, they explicitly address this and mention that they potentially over-estimate the unexplained pay gap, which I found positive. Another problem is the coarse-grained division into regions. The BFS uses the regions Geneva, Espace Middleland, Northwestern Switzerland, Zurich, East Switzerland, central Switzerland, and Tessin. While this division is fine for other surveys, it can be very poor for salary surveys as these regions contain places with very different salary structures. For instance, Kleinandelfingen has a much lower cost of living and lower salaries than Zurich and the same holds for Horrenbachbuchen (one of the poorest Swiss municipalities) and Bern, but they are in the same region (Zurich and Espace Middleland, respectively). This can potentially lead to biased results when estimating the gender pay gap (for instance, when one gender prefers to live in cities). A solution would be a regional division based on economic factors (e.g., Zurich, Zug, Basel, and Bern in one category, which are economically relatively similar) instead of geographic regions.

Furthermore, there may be a selection bias. Small companies with less than 20 employees are underrepresented with 23% of the investigated employees, although companies with less than 10 employees already employ roughly a quarter of Switzerland's workforce [2]. Although the survey is compulsory, the administrative effort to fill it out may be too large for many SMEs, but not for larger companies, which can simply create an ELM export in their HRMS (and have staff to do so). Therefore, the dataset may be skewed towards larger companies. This is partially addressed by the sampling weights, but there could still be a selection bias (for instance, the existence of a HR department might have an influence on the gender pay gap and also on the fact if a company responded to the survey). They also claim that only companies with at least three employees participate in the survey, which is not true. As the owner of a Swiss SME I participated multiple times in the study, even at times I had less than three employees (for instance in the 2016 wave that they used for their analysis). The BFS also does not mention this threshold and even explicitly says that owner-operated businesses without additional employees have to answer the survey [1].

2.2 Model

I found their choices of models and hyper-parameter selection methods (e.g., λ for LASSO) convincing and well justified. It is also good that they reported out-of-sample prediction power, as some of the models have the potential to overfit (with 615 control variables), even on large datasets.

For the exact matching estimator, I wonder what the impact of their usage of relatively coarse groups (for age and tenure) is. I could imagine that this degrades performance of the estimator and for a paper that analyzes the impact of methodology, I would have liked a more elaborate analysis. The problem cannot be completely circumvented (as they mentioned that too fine groups would result in empty strata), but different approaches could be tried, for instance resampling methods or generative machine learning models.

I also see room for improvement in the way they determine the order of the variables for imposing common support. They use the predictive power of a variable for a men's wage as the criteria for that. But in my opinion, one generally would like to include those variables first (i.e., impose common support on them early) that are able to explain differences between the genders well. These might overlap significantly with the ones that explain men's wages well, but there might also be differences. As an alternative, it would be possible to use the variables that reduce the predicted wage difference the most (according to some model, e.g. a simple linear one like they did for men's wages).

3 Further Work

It would be interesting to see which results also hold for smaller datasets. In many countries, companies are forced to analyze if there is an unexplained gender pay gap nowadays and might also over-estimate it for the reasons that they addressed, so this issue is highly relevant. However, some of the methods might lead to poorer results when the sample sizes are much smaller (only a few hundred employees). For those smaller datasets, the feature selection method could also matter much more than in the paper (where the difference between the full model and the one selected with LASSO was relatively small), so the analysis of different automated feature selection techniques could then also be a relevant addition.

Another analysis that I would find interesting is how the common support analysis differs per region/country and what the relationship between this analysis, the gender pay gap, and other equality measures is. There have been studies that there are less female STEM grads in countries with more gender equality [3], so I wonder if we would also see less common support in countries with more gender equality.

One very important methodological choice that they did not address is the dataset. Like I explained above, certain important variables may be missing and there may be certain biases. Analyzing the impact of this will be very hard, but one idea that comes to my mind is Monte-Carlo simulation. One could simulate the labour market (as close to the reality as possible, e.g. with women having more longer career breaks because of children) and incorporate stochasticity (for salary negotiations, luck, etc...) into it to get a ground-truth (with a configurable gender pay gap) dataset. This would allow to simulate different data collection / survey methods (e.g., incorporating the actual work experience or using the age and education as a proxy) and analyze how close these estimates come to the ground truth unexplained pay gap.

References

[1] Bundesamt für Statistik. *FAQ*. `https://www.bfs.admin.ch/bfs/de/home/statistiken/arbeit-erwerb/erhebungen/lse/faq.html`.

[2] *Statistik der Unternehmensstruktur (STATENT), Beschäftigte und Arbeitsstätten: Geodaten 2017*. 9526894. Bundesamt für Statistik (BFS) / BFS.

[3] Gijsbert Stoet and David C. Geary. “The Gender-Equality Paradox in Science, Technology, Engineering, and Mathematics Education”. In: *Psychological Science* 29.4 (Apr. 2018), pp. 581–593. ISSN: 0956-7976. DOI: `10.1177/0956797617741719`.

[4] Anthony Strittmatter and Conny Wunsch. *The Gender Pay Gap Revisited with Big Data: Do Methodological Choices Matter?* Tech. rep. 2021/05. Faculty of Business and Economics - University of Basel, Feb. 2021.