# Policies to Reduce Gender Inequality: Evidence from the U.S.

Fewo Nett

July 31, 2023

# Contents

# 1 Introduction

The proposed implementation of three labor market policies gave rise to vigorous political discussion in recent weeks. Namely, the firm internal publication of wages, the encouragement of salary negotiations, and the incentivization of companies to offer childcare services are at the center of discussion. The implementation of these exact policies in several US states presents an excellent opportunity to objectively analyze their effectiveness. This report will use a novel dataset on gender inequality to assess whether the reforms were successful in reducing gender inequality across wages and employment. A special emphasis will be given to the heterogeneity in effects across different groups in the sample. Furthermore, the transferability of the results to Germany will be discussed.

The report is structured as follows: In the data section, the contents of the data set are briefly outlined and the first difference transformation of the data set is explained. The analysis section covers descriptive and causal effect analysis of the reforms on both wages and employment. Section 4 explores threats to identification and the external validity of the results specifically in Germany.

# 2 Data

The baseline data contains two observations each for 1047 individuals in total. It offers a rich set of covariates: Basic information such as age, ethnicity, and gender is covered as well as measures of education, socioeconomic background, IQ, occupational knowledge, experience and more.

The measures of father and mother education were excluded for most of the analysis because they contain a high number of NAs, which leads to the loss of a large number of rows in subsequent data transformation and analysis. Sensitivity tests show that results are similar when including the variables. Since the individual's education as well as many other measures concerning family background are observed, it seems unlikely these variables carry a strong signal that other covariates cannot account for.

In order to use state of the art machine learning algorithms for variable selection and analysis of heterogeneous treatment effects, the data was transformed from a panel into a cross section using a first difference approach: The data was split into the 2005 and 2010 observations for all individuals. Subsequently for each individual the difference between years was calculated for each variable by subtracting the values of 2005 from the values of 2010. This was done for all 17 remaining variables including the outcomes wage and employment. To account for the loss of information due to time invariant variables and the potential relevance of the absolute values for treatment effect heterogeneity the covariates of the year 2005 were additionally included. This way the effect of level and change in each variables can be captured in the model.

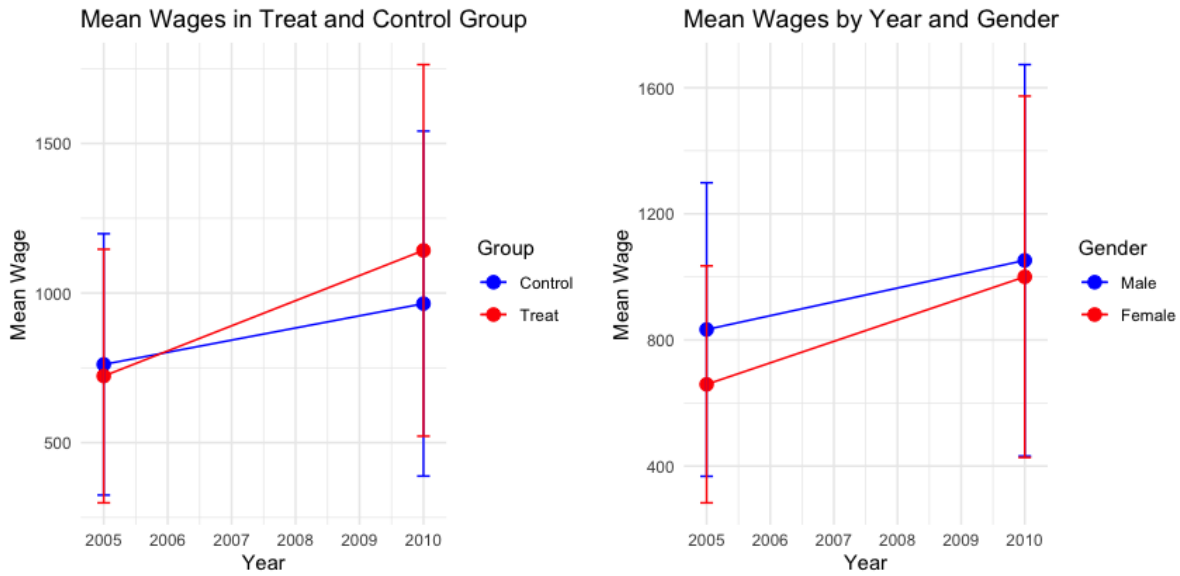


Figure 1: Mean wages over time

# 3 Analysis

## 3.1 Wages

### 3.1.1 Descriptive analysis

The gender gap in the entire sample has narrowed over time. The difference in average earnings between male and female observations in 2005 was 174$ and reduced to 52$ in 2010. Figure 1 presents the mean wages of treated and untreated observations over time on the left side and the mean wage by year and gender on the right side. Standard deviations are plotted as error bars. On the right hand side the reduction in the gender pay gap is clearly visible. Over time, wage between males and females appear to converge. Contrarily, the mean wages between the group affected by the reform and the control group are almost identical at baseline, but separate in 2010 as the reform group shows a stronger increase in wages. In the following analysis I will try to investigate whether the reduction of the gender pay gap and the increase in overall wages in the affected firms are caused by the reform, or whether it is a mere correlation.

### 3.1.2 Methodology

In a first step, the methods used to unpack the causal effects of the reforms will be discussed. As previously mentioned, the data was transformed into a cross section to allow for the usage of dimension reduction techniques. Specifically, double lasso (partialling out) estimators implemented in the hdm package (Chernozhukov et al. [2016]) were used to get a first overview of treatment effects. The partialling out double lasso method is specifically useful since it offers a neat way of dimension reduction to avoid overfitting the data, but simultaneously avoids regularization bias by performing the actual inference in a separate regression on the residuals. In order to investigate heterogeneity in treatment effects, the treatment indicator was interacted with all control variables. Subsequently, these interaction terms were used as treatment indicators in following regressions. This method offers a convenient way to explore the potential of variables to drive heterogeneity. However, the OLS estimation of residuals in the last step is not inherently robust to treatment effect heterogeneity. Therefore, the coefficients of these interactions should be considered cautiously. In the context of this analysis, they were used as an indicator, whether specific variables should be investigated further in their potential to drive heterogeneity using an augmented inverse probability weighting (AIPW) estimator which is heterogeneity robust. Specifically, causal forests (Athey et al. [2019]) implemented in the grf R package (Tibshirani et al. [2023]) were used in order ro confirm and further investigate evidence of heterogeneity in treatment indicators that was found using double lasso. Besides robustness to treatment effect heterogeneity, the nonparametric nature of causal forests is another upside of this estimation technique. The forest does not assume any underlying functional form of CATEs and is capable of capturing the effects of interactions between variables without explicitly including them in the model: A forest splitting first on age and then on education will capture the joint effects of the two variables in the resulting node.

The necessary assumptions for unbiased estimation of average treatment effects (ATE) and conditional average treatment effects (CATE) with such an estimator are common support and conditional independence. To justify the assumption of common support, a random forest was used to predict propensity scores on the available covariates. The conditional independence assumption leans on the correct selection of covariates. Given the sample size, one cannot hope to get accurate insights in heterogeneity when including all variables in the causal forest. Although the first difference transformation of the data helps to tackle issues of endogeneity, a careful selection of covariates that account for the remaining variation is still important to avoid biased estimates. The partialling out estimator was once again used to select variables for the causal forest: Variables were included if they had a non zero coefficient in one of the two lasso estimators (treatment indicator and wage difference as outcome variables) or if they significantly drove heterogeneity in interaction with treatment in the outcome variable.

### 3.1.3 Results

Table 1: Estimates and significance testing of the effect of target variables

| | Estimate | Std. Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| **Full sample** | | | | |
| treat | 12.53 | 20.69 | 0.605 | 0.545 |
| treat*fem | 302.44 | 26.92 | 11.234 | <2e-16 *** |
| **Full sample** | | | | |
| treat | -54.13 | 35.67 | -1.518 | 0.129 |
| treat*emp | 263.58 | 38.36 | 6.871 | 6.36e-12 *** |

The average treatment effect on wages in the entire population amounts to roughly 180$ across various specifications and methods. However, this number is misleading as treatment effects vary drastically across groups in the sample. In a first step, it will therefore be established which subgroups of the population are affected by treatment. In table 1 the results of two different partialling out double lasso regressions are presented. The estimates show the impact on the wage difference between 2005 and 2010. Panel one presents a strong significant coefficient on the interaction term of treatment and female, while the treatment indicator itself which can be interpreted as the effect on males does not show significant effects. Similarly, panel 2 demonstrates that only employed observations have a significant positive effect, while unemployed ones show no significant effects at all. Both these results were confirmed by running regression exclusively on male and untreated observations and no treatment effects were found. Importantly, the estimates are robust to including squared and interaction terms of all covariates and can also be replicated in a causal forest (as shown in code). This leads to the first main result: Only employed female individuals appear to be affected by treatment. This is important because it implies that all gains in wages for female observations can be directly interpreted as reducing the gender wage gap. Estimating effects for female observations with a causal forest yields a significant average treatment effect of approximately 320$, which translates into a reduction of the gender pay gap by the same amount. These insights allow to focus on the subgroup of employed females for the further analysis of heterogeneous treatment effects.

Next, the differences in treatment effects within the employed female populations will be inspected more closely using a causal forest. As an AIPW estimator, it needs estimates of the propensity scores as an input parameter. This lead to an interesting discovery along the way.

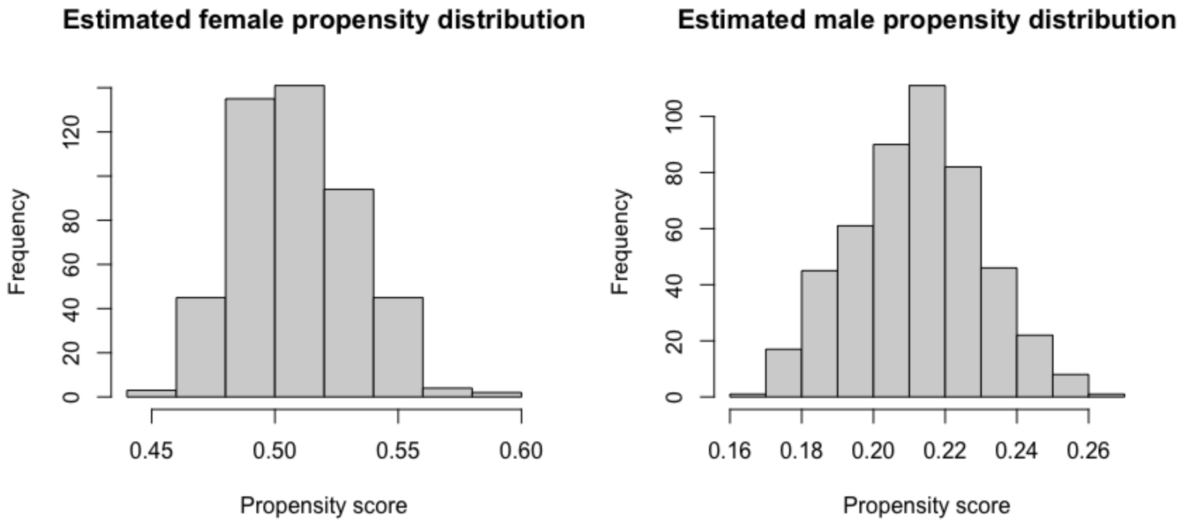


Figure 2: Propensity estimates by gender

Figure 2 presents random forest predictions of the propensity score distribution by gender. The propensity of male and female observations appears to be distinctly different: The female probability of being treated appears to be roughly distributed normally around 0.5, while the male probability is distributed around 0.2. No other variables appear to significantly drive heterogeneity. This does raise question regarding selection into treatment which will be discussed later on.

As explained above, the variable selection for the forest was performed via double lasso: Variables that have a non zero coefficient in one of the two lasso regressions (for outcome and treatment) were also included in the forest. Additionally, variables that were significant as treatment indicators when interacted with the treatment variable in a double lasso specification including all interactions and squared terms, were also selected.

Table 2: Best linear projection of the conditional average treatment effect. Confidence intervals are cluster- and heteroskedasticity-robust (HC3)

| | Estimate | Std. Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| (Intercept) | 349.770217 | 16.003578 | 21.8558 | < 2.2e-16 *** |
| educ | -38.339323 | 9.807631 | -3.9091 | 0.0001085 *** |
| age | -32.625537 | 5.827407 | -5.5986 | 3.985e-08 *** |
| urban | 62.613038 | 37.046242 | 1.6901 | 0.0917703 . |
| IQ | -3.427585 | 1.368944 | -2.5038 | 0.0126776 * |
| wage | 0.027223 | 0.087039 | 0.3128 | 0.7546223 |

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Table 2 presents the best linear projection of the conditional average treatment effects on the selected covariates. The significant intercept of 350 can be interpreted as the estimated average treatment effect for the employed female observations. Interestingly, years of education and age have a highly significant negative effect on the increase in wages due to treatment. Additionally, treatment effect intensity, appears to be weakly decreasing in IQ. This implies that less educated, young, and low IQ females experience a stronger increase in wages. The interpretation of the coefficients is straightforward. One year of additional education reduces the treatment effect by 39$ on average. These results are robust across several specifications of the causal forest including two different sets of covariates. However, there is no reason to assume that conditional average treatment effects are linear. Therefore, the conditional average treatment effects were estimated at different points of their respective distribution, to develop an intuition for their form. Figure 3 plots these estimates. On the left hand side the CATE are displayed at different percentiles of age. Standard errors are plotted around the estimates. They do appear to be broadly consistent with a downward sloping linear trend. On the right hand side the CATE are plotted against the respective number of education years. Besides the leftmost estimate the values appear to be broadly consistent with a downward sloping trend as well.

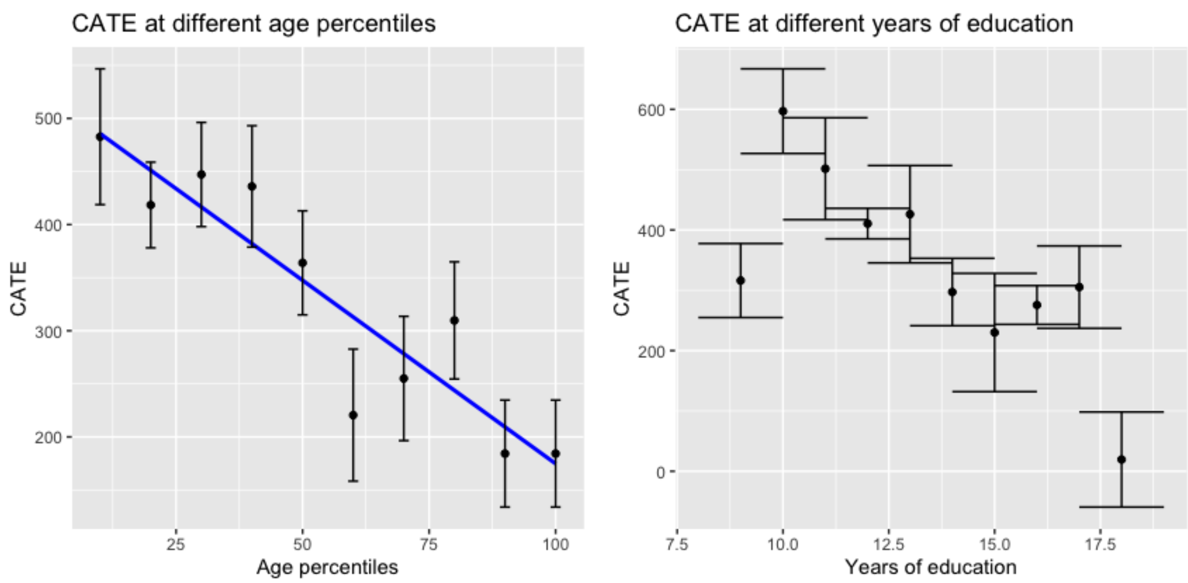


Figure 3: Propensity estimates by gender

## 3.2 Employment

There is little variation in the data in terms of employment. Between the years of 2005 and 2010 25 individuals went from being unemployed to being employed and 4 observations went from being employed to unemployed. The remaining 1018 observations did not change their employment status. This minimal amount of variation renders causal inference extremely difficult.

Attempts to estimate effects the impact of treatment on the probability of gaining or losing a job using a logistic double lasso model lead to no significant results. However , it is important to note that this does not mean that there are no effects of the reforms on employment. Given the low number of observations that switch employment status, even a strong treatment effect could go unnoticed. The sample is simply not large enough to allow for inference.

# 4 Discussion

## 4.1 Threats to identification

While the presented evidence is strong, there are some potential threats to causal inference within the sample. The first difference approach allows to address concerns of time constant unobservable characteristics and linear trends over time biasing the results, however it can not account for nonlinear trends in time, or differing time trends between control and treatment. This is problematic especially due to the financial crisis that took place around the same time and could expose the treatment and control groups to differing shocks. If such shocks cannot be accounted for with the observable variables, conditional independence is violated.

Another potential issue is hinted at by the big difference in propensities between male and female observations. This discrepancy points towards a selection into treatment. It is crucial to understand the difference between companies to which the reform applies and to which it does not to evaluate whether the observable variables can account for these differences. Once again, this is only problematic if the underlying variables driving the selection are not accounted for in the dataset. Out of all observables, the female indicator is the only strong predictor of selection into treatment.

While these concerns are valid, I believe the richness of the covariates mitigates them substantially: Characteristics like ability, that are typically unobservable in economic settings can plausibly be controlled fore with measures of IQ and occupational expertise.

## 4.2 Transferability of the results to Germany

In general, the results from the United States can not be directly transferred to other countries. The differences in institutions, culture, workforce characteristics, and more could result in differing effects of the policies in Germany.

Special regards should be given to the variables that have been identified to drive heterogeneity in the predetermined analysis when assessing how these policies would affect gender inequality in Germany. Specifically, the share of employed females in Germany should be considered as those individuals are most likely to benefit strongly from treatment according to the analysis. A higher share of females in the workforce compared to the US would therefore imply stronger average effects overall. Additionally, other variables that appear to drive heterogeneity should be considered. An analysis of age and education in the German female workforce is warranted to investigate possible effects.

Attention should also be paid to the baseline compliance to the policies: If most companies in Germany already encourage salary negotiations, offer childcare, and publish salaries internally while most companies in the US did not, there likely will be a smaller effect in Germany. This additionally poses the problem that compliance rates for each individual policy of the three might differ individually between countries ex ante. Since all policies were introduced simultaneously, it is not possible to differentiate their effects.

Further analysis and data is necessary to evaluate to which degree the results are transferable to Germany.

# 5 Conclusion

The introduced policies did have a strong positive causal effect on wages of employed females in the US. At the same time, unemployed and male individuals were likely unaffected by treatment. The gender wage gap reduced by approximately 320$ as a result. For employed female individuals the policies increased wages by 350$ on average. Within this subgroup treatment effects appear to be heterogeneous: On average younger and lower educated individuals experienced a greater increase in wages due to the treatment. The data also indicates that the effect of the policies on wages is weakly decreasing in IQ points, but the effect is less significant and further analysis is needed.

With regards to employment the report remains inconclusive: There is very little variation in the dataset and an effect can therefore neither be confirmed nor ruled out.

The external validity of the results within Germany is also unclear. To estimate the effect intensity careful analysis is needed to investigate whether Germany is similar to the US across relevant metrics like baseline female employment, compliance to the policies, and relevant workforce characteristics. Given the significance of the results in the US however, it does seem likely that there would be some reduction in gender inequality if the policies were to be implemented in Germany.

# References

Susan Athey, Julie Tibshirani, and Stefan Wager. Generalized random forests. 2019.

Victor Chernozhukov, Chris Hansen, and Martin Spindler. hdm: High-dimensional metrics. *arXiv preprint arXiv:1608.00354*, 2016.

Julie Tibshirani, Susan Athey, Rina Friedberg, Vitor Hadad, David Hirshberg, Luke Miner, Erik Sverdrup, Stefan Wager, Marvin Wright, and Maintainer Julie Tibshirani. Package 'grf', 2023.

I confirm that this report is based on my own work. In preparing this report, I have not received any help from another human nor have I discussed any aspects of the empirical project with others.