26th International Congress on Modelling and Simulation, Adelaide, SA, Australia, 30 Nov–4 Dec 2025
mssanz.org.au/modsim2025

Gender wage gap analysis using machine learning methods

Rob Dunne [a]

[a]*Data61, CSIRO Level 5, 13 Garden Street Eveleigh NSW 2015*
Email: rob.dunne@csiro.au

Abstract: We examine how modern machine learning methods, particularly Generalized Random Forests (GRF), can enhance our understanding of economic and social data. We demonstrate this with data on the gender wage gap using Public Use Microdata Area data from the United States Census Bureau. The gender wage gap remains a persistent issue, with women earning approximately 20% less than men as of 2024. Traditional econometric approaches have provided valuable insights but may miss complex interactions and heterogeneous effects that machine learning methods can help uncover. Our analysis partitions the workforce into three segments based on propensity scores: female-dominated ($e < 0.1$), mixed workforce amenable to causal analysis ($0.1 \leq e \leq 0.9$), and male-dominated ($e > 0.9$) sectors. We find that the gender wage gap varies significantly across demographic characteristics, with particularly pronounced effects related to age, children, and marital status. We do not find occupational segregation to be a large factor in the gender wage gap.

Keywords: *Gender wage gap, causal inference, machine learning, generalized random forests, labor economics*

1 INTRODUCTION

The gender wage gap remains a persistent issue in Australia, with women earning approximately 20% less than men as of 2024. Traditional econometric approaches have provided valuable insights but may miss complex interactions and heterogeneous effects that machine learning methods can help uncover.

We consider some of the approaches in the literature, and the data sets they are based on before introducing the Generalized Random Forests (GRF) method.

1.1 Decomposition of factors

Standard methods for estimating the relative importance of various causes of the gender wage gap include:

- Walby and Olsen simulation technique
- Oaxaca-Blinder decomposition
- Juhn-Murphy-Pierce decomposition

An example of the Walby-Olsen approach is the "She's Price(d)less" report prepared by KPMG in collaboration with Diversity Council Australia (DCA) and the Workplace Gender Equality Agency (WGEA) (KPMG, 2022). The report identifies key drivers of Australia's gender pay gap as:

- **gender discrimination** (36% of the gap, down from 39% in 2017) - this refers to systemic or overt discrimination that cannot be explained by factors such as education, experience, etc.
- **care, family responsibilities and workforce participation** (33% of the gap, down from 39% in 2017)
  - years not working due to interruptions (20%)
  - part-time employment (11%)
  - unpaid care and work (2%)
- **Gender segregation in job type** (24% of the gap, up from 17% in 2017)
  - industrial segregation (20%, up significantly from 9% in 2017)
  - occupational segregation (4%, down from 8% in 2017)

1.2 A psychology approach

A recent paper (Flinn et al., 2024) argues that the gender wage gap is primarily driven by differential returns to personality traits rather than differences in the traits themselves. Their approach incorporated Big Five personality traits into a structural job search model.

They develop a "job search, matching, and bargaining model" where personality traits affect labor market outcomes through four distinct channels:

- Worker productivity;
- Job offer arrival rates;
- Job dissolution rates;
- Bargaining power.

Unlike previous studies with gender-only heterogeneity, all job search parameters are index functions of individual characteristics including continuous personality measures. They use a large longitudinal data set with detailed Big Five personality assessments and job transition histories (Flinn et al., 2024). They fit their model by maximum likelihood and then explore the model via simulations with counterfactual interventions.

They find that men and women have different traits but also receive different labor market rewards for identical traits, particularly:

- Conscientiousness: Positive return for men, negative for women
- Agreeableness: Negative for both, but more negative for women

Unlike Oaxaca-Blinder or Walby-Olsen decompositions, this approach identifies causal mechanisms through which personality creates wage gaps, enabling policy-relevant counterfactual analysis. They argue that the wage gap stems from discriminatory valuation of personality traits rather than human capital, productivity, or personality differences per se. If women received the same returns to their personality traits as men, the entire wage gap would disappear.

1.3 Claudia Goldin's work

Claudia Goldin won the 2023 Nobel Prize in Economics for her work on understanding the gender pay gap. Goldin's work (see "Family and Careers" Goldin (2021)) puts the gender wage discrimination in the US into a historical context. She follows the careers of successive cohorts of college graduates, which she characterizes as "career or family", "career then family" and "career and family". While female college graduates started as a small group in 1900, it is now almost 1/2 of any emerging cohort of women.

While she documents the historical barriers that women faced in employment, she argues that discrimination by policy, management or fellow workers is, at most, responsible for a small part of the gender wage gap. Likewise she argues that neither unwillingness to negotiate or hold out for a higher remuneration, or "occupational segregation" is more than a minor cause. She notes that many large corporations have staff, policies and departments charged with "de-biasing" the workplace; however, she argues that it is not tackling the structural problem.

She locates the problem in what she calls "greedy work":

- employers pay a large premium for working longer hours and being on call at any time;
- any interruption to work history is heavily penalized. In a cohort of MBA graduates 7 years after graduation, women had taken 0.37 of a year off, men had taken 0.075. This, she argues, had a significant impact on career advancement;
- tasks where a professional organisation has regularized the hours will have a lower gender wage gap;
- veterinary work in the US is now largely organized into chains. This leads to a situation where fewer veterinarians are now small business owners and work hours are more predictable. As a consequence, veterinary practice in the US is now a female dominated profession and has a lower gender wage gap;
- areas where employees have been successful in demanding more predictability and flexibility, and where it is commonly recognized that work can be shared and that even highly skilled employees are interchangeable, will have lower gender wage gaps.

Goldin argues that the gender wage gap is highest in the top ranks of the most highly paid professions, and is highest in professions where there is a steep ladder for success. Academia in the US has been characterized as "move up or move out," with an average of 7 years to get tenure. This sort of pressure in academia, finance and legal services may work against women due to greater demands on their time. See also Doan et al. (2025) for an analysis of this in the Australian context.

2 DATA

We use Public Use Microdata Area (PUMS) data from the 2018 American Community Survey by the US Census Bureau (ACS, 2018). The analysis dataset contains 644,925 observations of individuals under 54 years of age. Each observation is characterized by a triple $\{X_i, W_i, Y_i\}$ where:

- X_i represents a vector of covariates for individual i including age, race, education level, marital status, number of children, occupation, industry, and geographic region;
- $W_i \in \{0, 1\}$ where 0 represents women and 1 represents men (following causal inference terminology where this is the "treatment" indicator);
- Y_i is the observed log wage outcome.

2.1 Generalized Random Forest Methodology

Our approach uses Generalized Random Forests (GRF) (Athey et al., 2019), which extends Breiman's random forests to estimate any quantity of interest identified as the solution to a set of local moment equations. For causal inference, we seek to estimate the conditional average treatment effect:

$\tau(x) = \mathbb{E}[Y_i(1) - Y_i(0)|X_i = x]$

where $Y_i(1)$ and $Y_i(0)$ are potential outcomes under treatment and control, respectively. In addition we define:

- $\mu(x, w) := E[Y_i|X_i = x, W_i = w]$
- $e(X_i) := \mathrm{P}[W_i = 1|X_i]$, the propensity score.

The GRF Algorithm. The GRF algorithm differs from standard random forests in its approach to recursive partitioning. Instead of splitting nodes to minimize prediction error, GRF splits to maximize heterogeneity in the parameter of interest $\tau(x)$. The algorithm proceeds as follows:

1. **Gradient-based splitting**: For each potential split, gradient-based approximations to the treatment effect estimating equation are computed. This involves:
   - Computing pseudo-outcomes using gradients of the moment condition

$$\mathbb{E}[\psi_{\tau(x),\nu(x)}(O_i)|X_i = x] = 0$$

   - Using these pseudo-outcomes to guide CART-style recursive partitioning
2. **Honest estimation**: Use sample splitting where one half of the data is used for tree construction and the other half for parameter estimation within each leaf, preventing overfitting. This differs from standard random forest implementations
3. **Subsampling**: Each tree is grown on a random subsample of the data to ensure stability and reduce correlation between trees.
4. **Adaptive weighting**: The final estimate $\hat{\tau}(x)$ uses forest-based weights $\alpha_i(x)$ that reflect how often training example i falls in the same leaf as test point x across all trees.

AIPW Estimation. The algorithm implements the Augmented Inverse Propensity Weighted (AIPW) estimator, which is doubly robust and efficient under unconfoundedness. The AIPW estimator for the average treatment effect is:

$$\begin{aligned}\hat{\tau}^{AIPW} := \frac{1}{n}\sum_{i=1}^{n} & \hat{\mu}^{-i}(X_i,1) - \hat{\mu}^{-i}(X_i,0) \\ & + \frac{W_i}{\hat{e}^{-i}(X_i)}\left(Y_i - \hat{\mu}^{-i}(X_i,1)\right) - \frac{1-W_i}{1-\hat{e}^{-i}(X_i)}\left(Y_i - \hat{\mu}^{-i}(X_i,0)\right)\end{aligned}$$

where $\hat{\mu}^{-i}(X_i, w)$ and $\hat{e}^{-i}(X_i)$ are leave-one-out estimates of the outcome regression and propensity score, respectively.

Key Assumptions. Causal identification relies on three key assumptions:

1. **Unconfoundedness**: $Y(1), Y(0) \perp W|X$, meaning potential outcomes are independent of treatment assignment conditional on observed covariates
2. **Overlap**: $0 < e(x) < 1$ for all x in the support of X, ensuring both treatment and control observations exist in all regions of covariate space
3. **SUTVA**: The Stable Unit Treatment Value Assumption, requiring no interference between units and treatment variation irrelevance

We partition the data into three segments to address the overlap assumption, as the extreme segments ($e < 0.1$ and $e > 0.9$) violate this condition, see Figure 1. We have:

- the female dominated segment ($e(X_i) < 0.1$): 40,063 observations;
- the part of the workforce amenable to causal analysis ($0.1 \le e(X_i) \le 0.9$): 644,925 observations;
- the male dominated segment ($e(X_i) > 0.9$): 76,792 observations.

We will have to analyse the $\{$ low, high$\}$ propensity segments separately.

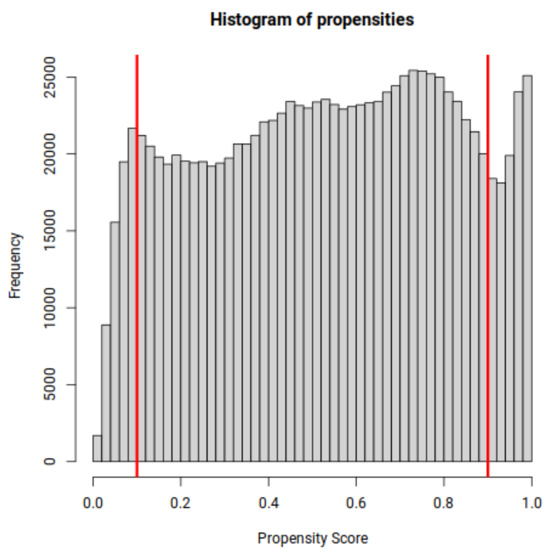

Figure 1. Propensity Scores, vertical lines at 0.1 and 0.9 indicate where the propensity was trimmed to ensure that the overlap condition is met. This gives three segments to be analysed, namely the $\{$low, mid, high$\}$ propensities.

2.2 Diagnostics

We start by fitting the causal forest model to estimate τ and running some diagnostic tests. We start with a calibration test. We fit the linear model:

$$Y_i - \hat{m}(X_i) = \alpha\bar{\tau}(W_i - \hat{e}(X_i)) + \beta(\hat{\tau}(X_i) - \bar{\tau})(W_i - \hat{e}(X_i)) + \epsilon_i$$

The coefficients have the following interpretations:

- α (mean forest prediction): Should be close to 1 if the average predicted outcome equals the average observed outcome. The coefficient is 1.004, which is good (p-value $< 2.2 \times 10^{-16}$);
- β (differential forest prediction): Should be close to 1 if the forest is capturing true effect heterogeneity. The coefficient is 1.037, which is good (p-value $< 2.2 \times 10^{-16}$).

We can assess the balance of the variables between male and female before and after propensity adjustment. We plot the absolute standardized mean difference (ASMD) of X between treated and untreated individuals for the raw data and also for the propensity weighted data. A visual inspection shows the variables are well balanced after the adjustment (Supplementary file 3).

Unlike balance and overlap, which can be tested using observed data, unconfoundedness (conditional independence) is fundamentally untestable because it involves potential outcomes that are not observed. The unconfoundedness assumption states that:

$$Y(1), Y(0) \perp W|X$$

In words, the potential outcomes are independent of the treatment assignment conditional on observed covariates X. The assumption is basically that we have measured the correct variables.

3 RESULTS

3.1 Variable importance and Best Linear Projection

The causal forest produces a measure of variable importance as shown in Table 1. It is the results of a highly non-linear process so the rankings are informative but the numerical values are not easily interpretable. These can be compared with the best linear projections which provide simple, interpretable summaries of which variables drive variations in treatment effects. However there is a strong assumption that the treatment effect is linear in these covariates and that there are no interactions.

Putting these two sources of information together, we can conclude that marital status, age, and number of children are the main drivers of the gender wage gap. Years of education has an ameliorating effect on the wage gap (it has a negative coefficient).

Figure 2 shows the interactions of age and marital status, the two most significant drivers of the wage gap. We see that the never-married group has the smallest wage gap and that it is relatively flat across age. Conversly, the wage gap for the married cohort increases throughout their working life.

3.2 High and low propensity segments

Table 2 shows a comparison of the some variables across the low, mid and high propensity segments. The propensity measure deals with the coverage of the sample space, so these segments are regions where τ can not be estimated, which could be due to occupational segregation or any number of other causes. In addition, all these segments have in common is the propensity score, so they may not be homogeneous.

We note that the wage gap is lower for the low and high propensity segments than it is for the mid segment.

Figure 3 shows occupations, ranked by average wages. Several occupations that make up a high proportion of wage earners in the high and low propensity segments have been colored. Those that make up more than 7 of the respective segments have been labelled. It appears that the high and low propensity occupations cover a range of wages.

We can calculate a counterfactual scenario where we ask "What would women's average wage be if they had men's occupational distribution but kept their own within-occupation wage rates?". This suggests that only 11.5% of the gender wage gap is explained by occupational segregation.

Table 1. Best Linear Projection with Variable Importance (Ordered by Importance)

Variable	Importance	Rank	Estimate	Std. Error	t-value	Pr($>|t|$)
marital	0.640	1	0.0155	0.0009	17.43	$< 2 \times 10^{-16}$ ***
age	0.138	2	0.0035	0.0002	21.31	$< 2 \times 10^{-16}$ ***
education_level	0.096	3	-0.0066	0.0006	-11.86	$< 2 \times 10^{-16}$ ***
children	0.032	4	0.0239	0.0017	14.40	$< 2 \times 10^{-16}$ ***
occupation	0.031	5	-0.0001	0.00001	-4.84	1.3×10^{-6} ***
family_size	0.016	6	-0.0015	0.0011	-1.37	0.171
race	0.014	7	0.0083	0.0010	8.10	$< 2 \times 10^{-16}$ ***
english_level	0.011	8	-0.0115	0.0030	-3.77	1.6×10^{-4} ***
economic_region	0.009	9	-0.0028	0.0006	-4.79	1.7×10^{-6} ***
hispanic_origin	0.006	10	-0.0151	0.0045	-3.36	7.9×10^{-4} ***
nativity	0.005	11	-0.0046	0.0052	-0.88	0.379
citizenship	0.003	12	-0.0041	0.0018	-2.28	0.022 *
industry	0.002	13	-0.00001	0.00002	-0.59	0.556
(Intercept)	—	—	0.0917	0.0190	4.81	1.5×10^{-6} ***

Note: Variables ordered by causal forest variable importance scores. Importance scores represent weighted frequency of splits in the causal forest. Significance codes: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.
Interpretation: Positive estimates indicate variables associated with larger gender wage gaps; negative estimates indicate variables that reduce the gap.

Table 2. Demographic Comparison Across Propensity Score Segments

	Count		Distribution		Age		Education		Children		Wage Gap
Segment	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Gap
Low	2,089	37,974	5.2%	94.8%	38.1	39.2	18.7	18.4	0.86	0.93	0.16
Mid	330,231	314,694	51.2%	48.8%	37.1	37.0	19.1	19.4	0.82	0.76	0.19
High	74,260	2,532	96.7%	3.3%	37.4	37.7	16.5	17.3	1.07	0.97	0.13

Note: Sample restricted to individuals under 54 years old.
Segments: Low = female-dominated sectors; Mid = mixed workforce; High = male-dominated sectors.
Gap: Raw gender wage gap (%)

4 CONCLUSIONS

This paper demonstrates how machine learning methods, specifically Generalized Random Forests, can provide new insights into the gender wage gap. Our analysis reveals significant heterogeneity in treatment effects across different demographic groups and workplace characteristics.

Key findings include the observations that:

- marital status is the main predictor of the gender wage gap;
- the gap increases significantly with age for the married group but is much flatter for the never-married;
- the "motherhood penalty" is a strong predictor of the wage gap;
- education partially mitigates but does not eliminate the gender wage gap.

These findings are in line with Doan et al. (2025) and Goldin (2021) in that they locate the problem in the "time economy". They differ from WGEA reports (KPMG, 2022), reports from Jobs and Skills Australia (Jobs and Skills Australia, 2025), and public comments by Jobs and Skills Australia commissioner (Commins, 2025) in giving a lower importance to occupational segregation.

As reports from the Workplace Gender Equality Agency and Jobs and Skills Australia are likely to set the policy agenda, these findings have important policy implications. This analysis suggests that addressing occupational segregation without addressing within-occupation wage rates may address occupational shortages but will not address the gender pay gap.

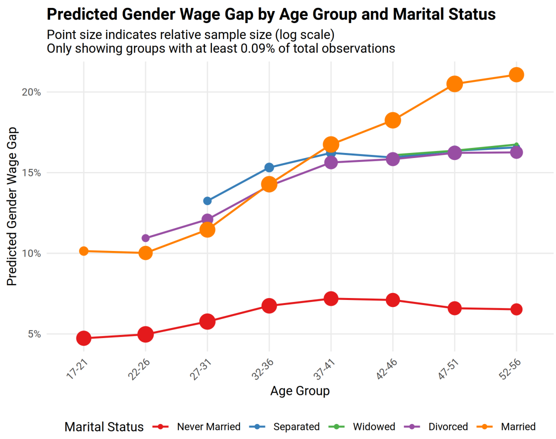

Figure 2. Age and Martial Status Interactions.

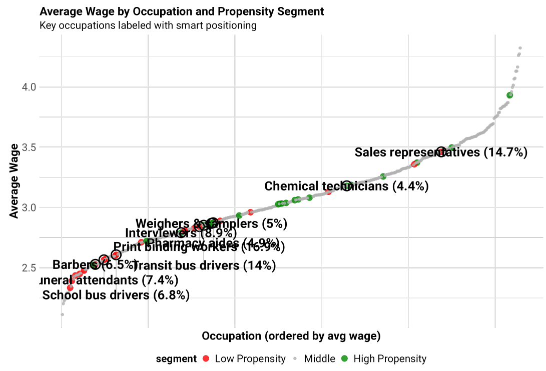

Figure 3. Occupations, ranked by average wages. Several occupations that make up a high proportion of wage earners in the high and low propensity segments have been colored. Those that make up more than 7 of the respective segments have been labelled.

Supplementary Material

All code for the analysis and the generation of the plots is available in

- Supplementary_1_The_PUBS_data_and_preprocessing.Rmd
- Supplementary_2_First_Look_at_the_Data.Rmd
- Supplementary_3_Analysis.Rmd
- Supplementary_4_segregated_workforce.Rmd

Acknowledgements

This research uses publicly available data from the US Census Bureau's Public Use Microdata Area (PUMS) dataset.

References

ACS (2018), Public use microdata area.
URL: *https://www.census.gov/content/dam/Census/data/developers/api-user-guide/api-guide.pd*

Athey Susan, Tibshirani Julie and Wager Stefan (2019), Generalized random forests, **47**(2), 1148–1178.
URL: *https://projecteuclid.org/journals/annals-of-statistics/volume-47/issue-2/Generalized-random-forests/10.1214/18-AOS1709.full*

Commins Patrick (2025), Data reveals worst sectors for gender pay inequality in Australia, *The Guardian* .

Doan Tinh, Leach Liana and Strazdins Lyndall (2025), The Economic Costs of Men's Long Work Hours for Women: Evidence on the Gender Wage Earnings Gap from Australia and Germany, *Social Indicators Research* .

Flinn Christopher J., Todd Petra and Zhang Weilong (2024), Labor Market Returns to Personality: A Job Search Approach to Understanding Gender Gaps, p. 734092.
URL: *https://www.journals.uchicago.edu/doi/10.1086/734092*

Goldin Claudia (2021), *Career and Family: Women's Century-Long Journey toward Equity*, Princeton University Press.
URL: *https://press.princeton.edu/books/hardcover/9780691201788/career-and-family?srsltid=AfmBOopbEjUtoYWSsbPAtulxRjIPmDmnAavlKWTH-p-wVgneaIIUnpaY*

Jobs and Skills Australia (2025), New perspectives on old problems: Gendered Jobs Work and Pay.
URL: *https://www.jobsandskills.gov.au/publications/new-perspectives-old-problems-gendered-jobs-work-and-pay*

KPMG (2022), She's price(d)less: The economics of the gender pay gap.
URL: *https://kpmg.com/content/dam/kpmg/au/pdf/2022/kpmg-shes-priced-less-2022.pdf*