Exploring the Gender Wage Gap within Industry

Miranda Lamb (mlamb36)
Last 3 digits: 881, mlamb36@gatech.edu
11/22/2025
ISYE 7406: Data Mining and Statistical Learning

Abstract

The gender wage gap has been a persistent and well-studied phenomenon in the United States- often attributed to various factors, including differences in industry. However, how the gender wage gap varies across industry has been explored less frequently. This study aims to explore the gender wage gap within industry, ultimately answering "how does the gender-wage gap vary across industries," by exploring industry impact on income. The study explores various models, and trains a random forest and linear regression model for reference, but ultimately relies on 2023 IPUMs data and a gamma GLM model with log link (with variables selected via forward-backward stepwise selection), to best fit the right-skewed income while preserving interpretability of the relationship. Predictive performance of the GLM and reference models were evaluated on a test set via RMSE, MAE, and r-squared. The GLM model performed comparably to the random forest and linear model, all with relatively weak predictive performance. To examine the relationship of interest, the coefficient on sex and interaction terms of sex and each industry were interpreted. Ultimately, there was a large difference in the magnitudes of gender wage gaps across industry, with 71.6% for the "Agriculture, Forestry, Fishing and Hunting, and Mining" industry, and 13.3% for the "Military" industry. These results can inform many industry decisions, and further research could explore these relationships in greater depth.

Introduction

Differences in financial power for men and women have been prevalent in the United States since its inception. Women's rights and earning power have increased significantly over the years, however the gap between men and women's earnings remains a prevalent field of study. According to the Pew Research Center, the gender wage gap has decreased slightly in the last 20 years, and in 2024 women earned, on average, 85% of what men earned. The contributors to this wage gap have been studied at length, with some studies finding evidence that the gap can be partially explained by career choices, educational differences, difference in lifestyle and expected familial responsibilities, and a myriad of other institutional and individual factors. Existing research qualifies the difference in salaries as a result of differences in occupation, however there is less research examining the wage gap between men and women within a given industry. Better understanding how the wage gap varies within each industry (rather than across industry) can help employers, policy makers, and other institutions make informed decisions to help close the wage gap, targeted at the industries that have the largest gaps.

This paper aims to explore the wage gap between men and women within a set of industries. The analysis is conducted using IPUMS data from 2023. The analysis is first explored through linear regression (with the response variable, income, transformed using the log function to account for the large right-skew of income data), however the bulk of the analysis is performed with a gamma log-link generalized linear model. This paper will yield a better understanding of GLM models, and will aim to answer the question "how does the gender wage gap vary within industry, based on the coefficient on gender and its interaction terms?"

Problem Statement and Data Sources
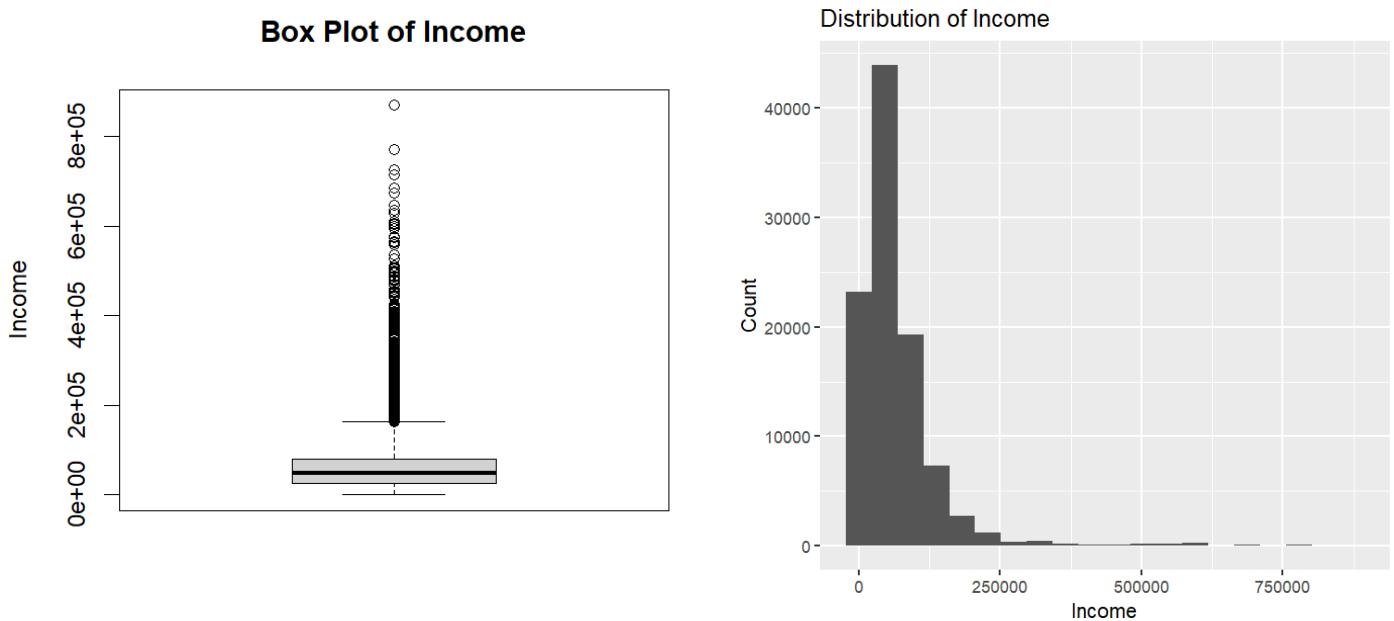
   This analysis is conducted using US Census data from usa.IPUMS.org from 2023. The dataset contains the following factors, and was filtered to only observations of those actively in the work force with employment. Each observation is one person in the year 2023.

- YEAR: Census year
- SEX
- AGE
- RACE
- EDUC: Educational attainment
- EMPSTAT: Employment status
- LABFORCE: Labor force status
- OCC: Occupation
- IND: Industry (coded into "Industry")
- INCWAGE: Wage and salary income
- PWSTATE2: Place of work: state
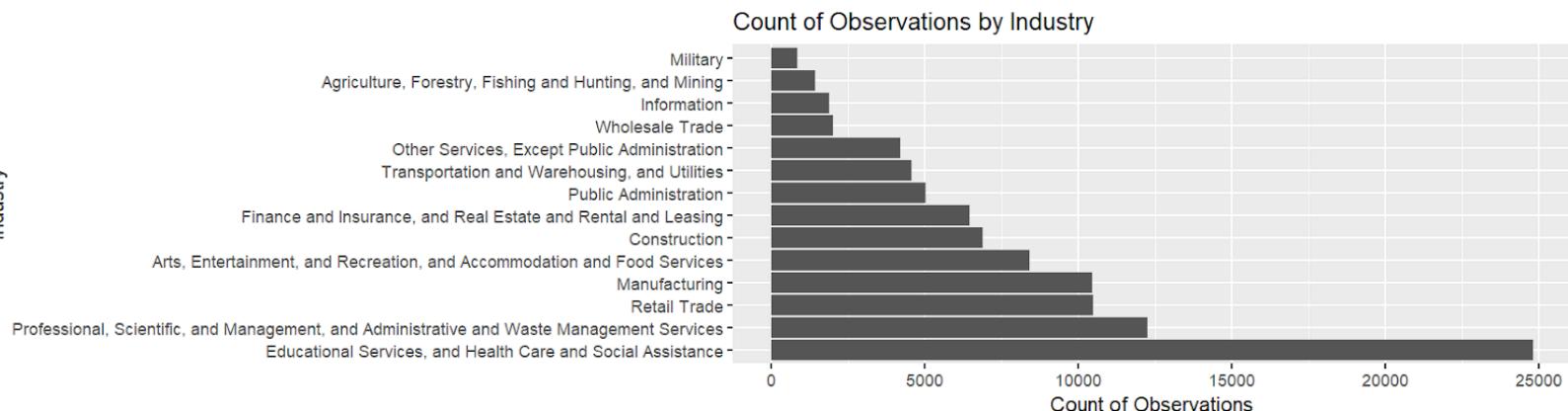
From this whole population set, a random sample of 100,000 observations was selected to serve as the dataset used for analysis. The main variables of interest are income (which in this data set is just wage and salary income), sex (coded to male/female), and industry, which is grouped into the following industries:

- Agriculture, Forestry, Fishing and Hunting, and Mining
- Arts, Entertainment, and Recreation, and Accommodation and Food Services
- Construction
- Educational Services, and Health Care and Social Assistance
- Finance and Insurance, and Real Estate and Rental and Leasing
- Information
- Manufacturing
- Military
- Other Services, Except Public Administration
- Professional, Scientific, and Management, and Administrative and Waste Management Services
- Public Administration
- Retail Trade
- Transportation and Warehousing, and Utilities
- Wholesale Trade
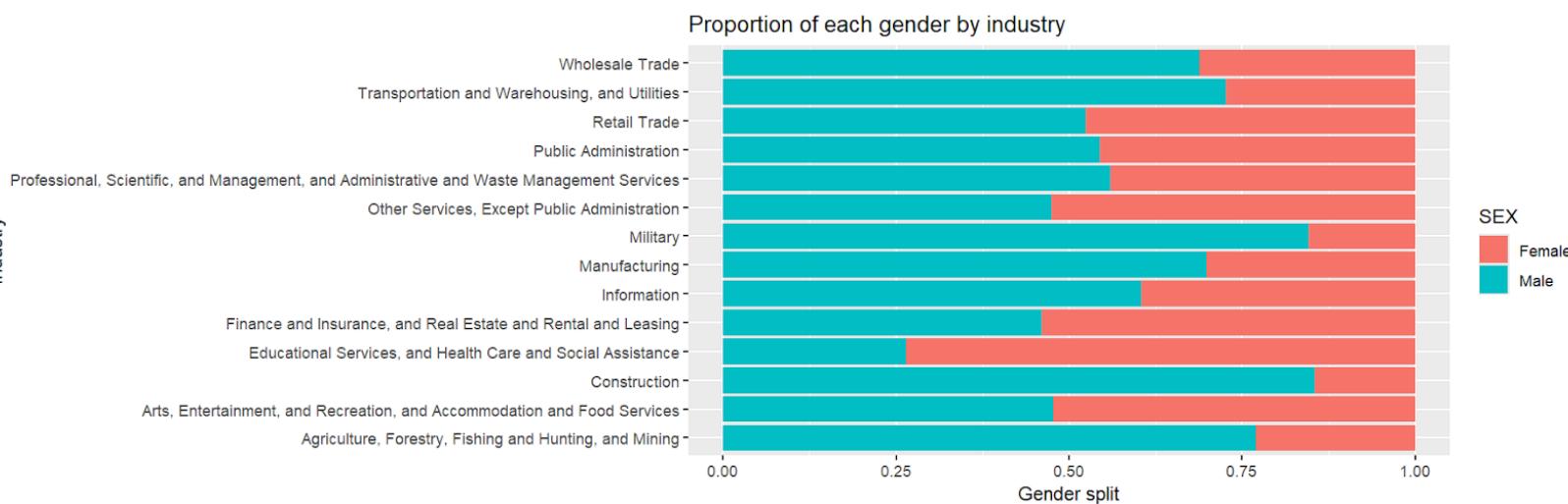
   To begin to explore the data set, the response variable of INCWAGE was plotted as a boxplot and a histogram. We can see from both of these plots, that the data is largely rightly skewed. This is to be expected, as the majority of people make lower incomes, with a long tail of the fewer individuals with higher incomes.

Box Plot of Income

Distribution of Income

Then, a plot was created to evaluate the number of observations for each industry in the data set There is a large sample size, however there is a smaller number of observations for some industries, like the Military.

Count of Observations by Industry

Next, a plot showing the proportion of each gender within each industry was created, to evaluate the balances in the data set. Each industry has a sufficient amount of observations of each gender due to the large sample size, however we can see that the military and construction industries are more than 75% male, and the Educational Services, Health Care, and Social Assistance industry is around 75% female.

Proportion of each gender by industry

Finally, a box plot was created of income for each gender (figure 5) where we can see the overall wage gap- as all quantiles for females are less than that of males.
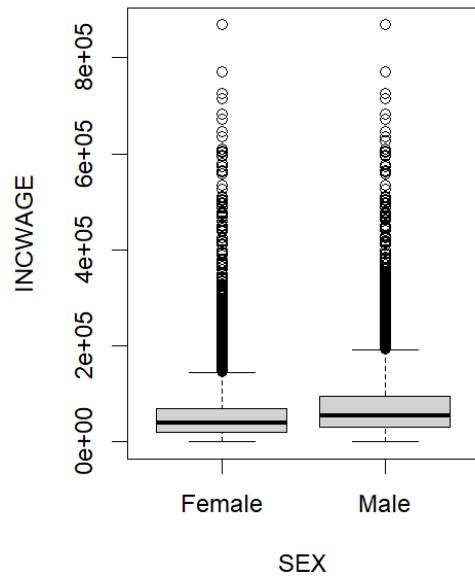
This data will be sufficient to explore our question- "how does the gender wage gap vary within each industry?"

Methodology

After initial data cleaning and exploratory data analysis, a simple linear regression was fit to the data to begin to explore the relationships, with the income variable as the response variable, an interaction term between SEX and Industry as the main predicting variable, and RACE, EDUC, and PWSTATE2 included as additional controls. Model assumptions were then tested with residual analysis, and assumptions were violated. So, the same linear regression was fit again, this time with the log of income as the response variable to account for the right-skew of the income distribution. Linear regression was explored first because of its interpretability. Model assumptions were then tested for the log-income model, and the corresponding residual plots can be seen in the appendix (figures 6-8). We can see that the constant variance assumption seems to be met, but by looking at the Q-Q plot, we see that the normality assumption is violated, and the log-transformed linear regression model is not a good fit to the data.

To better fit the data, while still preserving the interpretability of regression that will allow us to quantify the relationship between gender and industry on income, a gamma generalized linear model was fit to the data. The gamma model was chosen because it's well suited for continuous, positive, and right-skewed outcome distributions, like our income variable. Various link functions (inverse, identity) within the gamma model were explored, however ultimately the log-link was chosen as it enforces positive fitted values and produced stable model convergence (whereas the others did not, and also did not fit theoretically with the problem at hand). Regularization methods like Lasso and Ridge, combined with cross-validation to tune relevant hyperparameters were considered, however these methods optimize predictive accuracy rather than relationship derivation, and risk minimizing the true magnitude of the coefficient of interest. So, instead, variable selection was conducted using forward-backward stepwise selection to minimize AIC, to maintain the models estimations of each relationship while maintaining model parsimony. The best selection of variables was the full model, with the sex*industry interaction term, age, education, race, and state. Results of the stepwise variable selection comparisons can be seen in figure 9.

4

After arriving on the final log-link gamma model, the model was trained on a training set comprised of 70% of the data set. The Deviance residuals vs fitted values was plotted to evaluate model assumptions- this plot can be seen in figure 10. There is some slight fanning with a larger variance in the residuals for lower fitted values, however the residuals are mostly evenly scattered with a mean of zero, so we will consider the gamma model and the log link function a fair fit. Additionally, the model (with the interaction term removed, and sex and industry present independently) was evaluated for multicollinearity using the variance inflation factor. All VIF measures were less than two (results can be seen in figure 11), indicating multicollinearity is not a concern in this model.

Finally, after training and testing the fit of the final gamma log-link model, predictive performance was evaluated based on the remaining 30% of test data. Additionally, a log-linear regression and a random forest model were fitted on the training data and evaluated on the test data to serve as a performance comparison. Performance was evaluated on RMSE, MAE, and R-squared.

Analysis and Results

The predictive performance results of the log-link gamma model and the two comparison models, evaluated on the test data set are below.

Model	RMSE	MAE	R^2
Gamma log-link GLM	70,980.38	37,762.59	0.22
Linear Regression (log-linear)	74,139.17	36,447.08	0.22
Random Forest	69,679.92	37,531.4	0.23

We can see that Random Forest performed the best in terms of prediction accuracy. Although this model does not provide insight into the relationship at hand between income, sex, and industry, its often strong performance gives us a valuable comparison for our gamma log-link model's performance. We can see that the random forest model did not significantly outperform the gamma model. However, the mean of the income for the test data set is 67,426, so the RMSE values are relatively high for all models. This makes sense as the RMSE measure is sensitive to outliers/extreme values, which are present in income data. The MAE of our gamma model is 37,762.59. Taken together, the gamma model (as well as the other two comparison models) have relatively poor predictive performance, and only ~20% of the variance in income can be explained by our predicting variables (which is somewhat standard for large income data sets). However, the gamma model can still provide meaningful insight into the relationship of interest.

The full regression summary can be found in the appendix. The most relevant coefficients' percentage interpretations ((exp(coef. estimate) -1)*100) can be seen below.

| Variable	Percentage Change	Pr(>|t|)	Significance
(Intercept)		< 2e-16	***
SEXMale	72%	3.39e-16	***
IndustryArts, Entertainment, and Recreation, and Accommodation and Food Services	-27%	2.22e-07	***
IndustryConstruction	29%	0.000	***
IndustryEducational Services, and Health Care and Social Assistance	-4%	0.540	
IndustryFinance and Insurance, and Real Estate and Rental and Leasing	46%	7.83e-10	***
IndustryInformation	50%	2.39e-08	***
IndustryManufacturing	38%	2.65e-07	***
IndustryMilitary	27%	0.030	*
IndustryOther Services, Except Public Administration	-20%	0.001	***
IndustryProfessional, Scientific, and Management, and Administrative and Waste Management Services	34%	1.64e-06	***
IndustryPublic Administration	17%	0.014	*
IndustryRetail Trade	-14%	0.017	*
IndustryTransportation and Warehousing, and Utilities	18%	0.015	*
IndustryWholesale Trade	32%	0.000	***
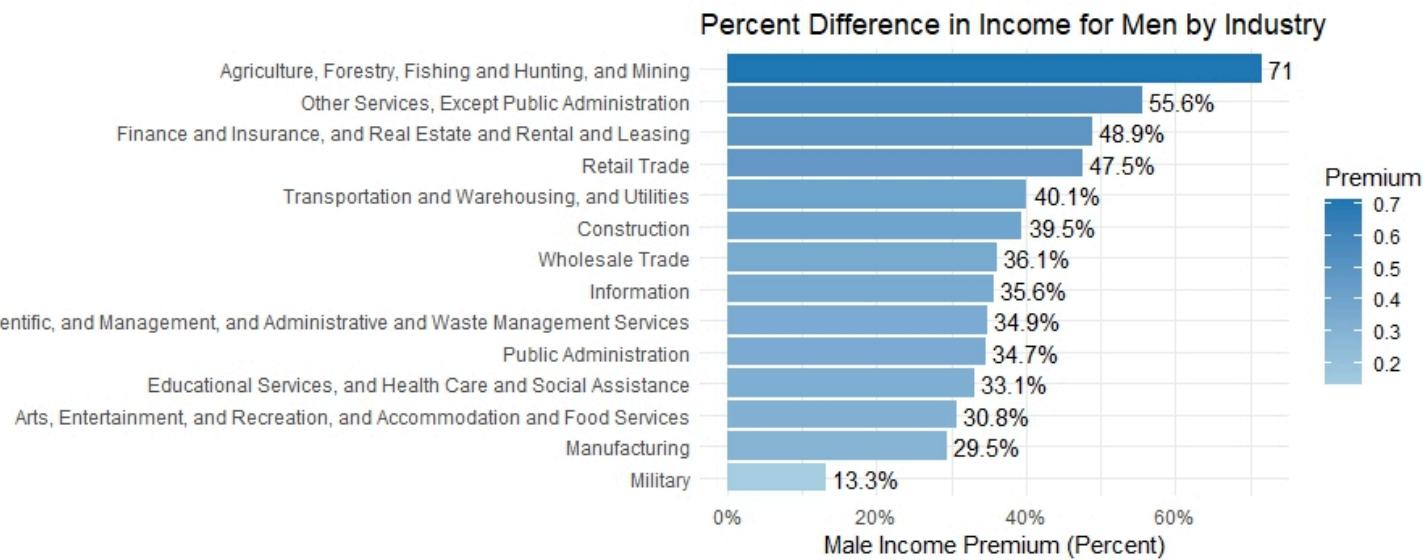
SEXMale:IndustryArts, Entertainment, and Recreation, and Accommodation and Food Services	-24%	0.000	***
SEXMale:IndustryConstruction	-19%	0.007	**
SEXMale:IndustryEducational Services, and Health Care and Social Assistance	-22%	0.000	***
SEXMale:IndustryFinance and Insurance, and Real Estate and Rental and Leasing	-13%	0.048	*
SEXMale:IndustryInformation	-21%	0.006	**
SEXMale:IndustryManufacturing	-25%	6.18e-05	***
SEXMale:IndustryMilitary	-34%	0.001	***
SEXMale:IndustryOther Services, Except Public Administration	-9%	0.186	
SEXMale:IndustryProfessional, Scientific, and Management, and Administrative and Waste Management Services	-21%	0.001	***
SEXMale:IndustryPublic Administration	-22%	0.001	***
SEXMale:IndustryRetail Trade	-14%	0.030	*
SEXMale:IndustryTransportation and Warehousing, and Utilities	-18%	0.007	**
SEXMale:IndustryWholesale Trade	-21%	0.006	**

With such a large sample size, significance values can become inflated- so rather than focusing on the p-value as the only measure of significance, we will also consider the magnitude of the effects taken with the p-values. All interpretations are relative to the baseline group, which is a white female in the "Agriculture, Forestry, Fishing and Hunting, and Mining" industry, with no schooling, and an age of zero.

First, looking at the "SEXMale" variable we can see that according to our model, males in the baseline industry are expected to have a 72% higher income than females, all else held equal. For reference, the overall ratio of mean men's income to mean female's income is 1.42-

meaning overall, agnostic of industry, we'd expect males' incomes to be around 42% higher than females' (not necessarily accounting for other variables). So, this 72% is relatively high- meaning the gender gap may be especially large in the baseline "Agriculture, Forestry, Fishing and Hunting, and Mining" industry. Looking at the industry variables, we can see that industry can have a large impact on income, which is to be expected. For example, a female in the "Information" industry is expected to have a 50% higher income, relative to one in the baseline "Agriculture, Forestry, Fishing and Hunting, and Mining" industry, all else held equal.

The focus of this analysis, however, relies on the sex/industry interaction terms. These values tell us how much the effect of being a male on income varies within industry, relative to men in the baseline industry. To fully interpret the difference in income between men and women for each industry, however, we needed to combine the coefficient on "SEXMale" with each interaction term's coefficient. So, for example, to answer "what is the expected difference in men and women's income in the construction industry, we would take 0.54 (the coefficient on SEXMale), added to -0.21 (the coefficient on SEXMale:IndustryConstruction), convert that value (.33) to the percent change: (exp(0.33)-1)*100 =39.1%. So, in the construction industry, males are expected to make 39.1% more than females. This calculation is shown for each industry below.

Percent Difference in Income for Men by Industry

Although all values are calculated, it should be noted that the coefficients on "SEXMale:IndustryFinance and Insurance, and Real Estate and Rental and Leasing", "SEXMale:IndustryOther Services, Except Public Administration", and "SEXMale:IndustryRetail Trade", are all not statistically significant at the 0.01 level- all other interaction terms and the SEXMale coefficient are statistically significant at (at least) the 0.01 level. So, for those results, we cannot conclude that the income effect of being a male in each of those industries is statistically different from a female in the baseline industry, at the 0.01 level. This could be due to relatively small sample sizes of males in each of those industries, or because a true lack of a relationship.

We can see that our model suggests that the largest wage gap (defined here as the difference in expected income for males relative to females) exists in the "Agriculture, Forestry, Fishing and Hunting, and Mining" industry at 71.6%, and the smallest wage gap exists in the "Military" industry at 13.3%. Most wage gaps are around 30-50% across other industries. In the contest of hypothesis testing, for all but 3 of the industries, we've rejected the null hypothesis that the coefficients on male and its interaction terms are zero, meaning that there are statistically significant differences in the wage gap across industries. These results ultimately answer our research question: "how does the gender wage gap vary within each industry?"

All results were found with https://www.r-project.org/, specifically the "ggplot2", "dplyr", "car", "MASS", and "RandomForest" packages.

Conclusions

In this paper, after building, validating, and assumptions testing several models, selecting the most appropriate (gamma-log) model, testing it, and finally interpreting the results, we have explored the question "how does the wage gap vary by industry". Limitations of this analysis include the fact that p-values were likely inflated as a result of the large sample size, some variables of interest were not statistically significant, and in some industries, there may have been a relatively small amount of observations for men and/or women, affecting how those relationships appear in the final result. Additionally, there may have been additional factors not accounted for, and the final model did not ultimately have strong predictive power- however still was useful in exploring relationships.

The results and interpretations found in this paper could be leveraged in many other contexts. For example: public policy aimed at reducing the gender wage gap might target the industries with the largest gap, or businesses in industries with larger gaps might take care to examine their own gender income equality. Future research areas could include what policies could help close the wage gap for each industry, which factors are more relevant to income inequality within each industry, and a deeper exploration into the gender imbalance within industry- perhaps using larger data sets, more segmented industries, and other controlling variables.

Appendix

Figure 6: Histogram of residuals from log-linear regression

Histogram of Residuals
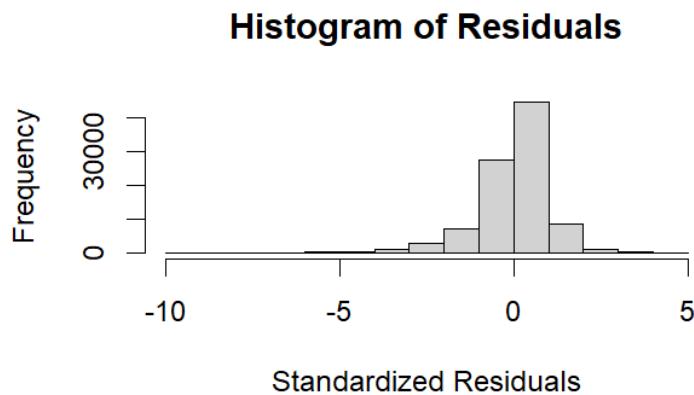
Figure 7: Normal Q-Q plot of log-linear regression residuals

Normal Q-Q Plot of Residuals
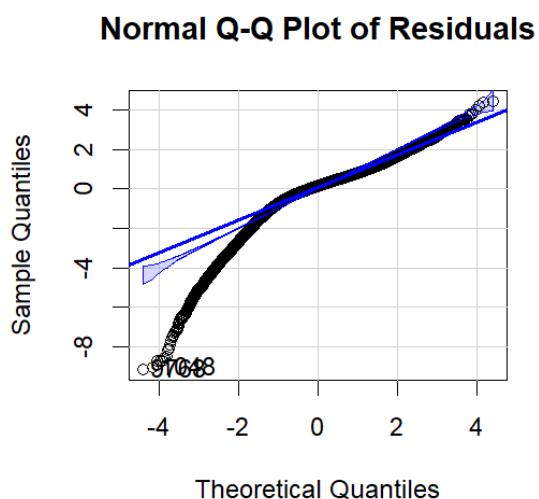
Figure 8: Fitted values vs standardized residuals for the log-linear model

Fitted values vs Standardized Residuals
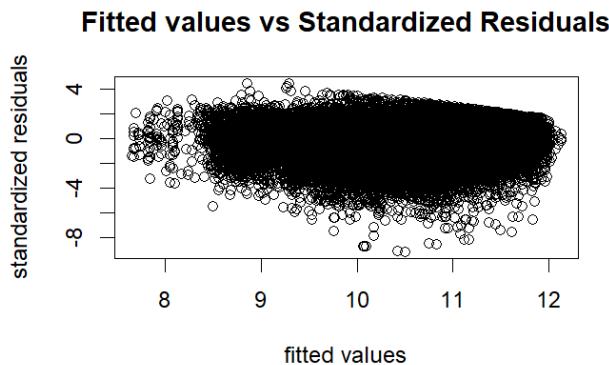
Figure 9: Results of the stepwise variable selection

	Df	Deviance	AIC
All variables		71505	2379165
SEX:Industry	13	71568	2379213
RACE	8	71761	2379448
PWSTATE2	56	73079	2380897
AGE	1	74713	2382922
EDUC	10	83741	2393485

Figure 10:

Residuals vs Fitted
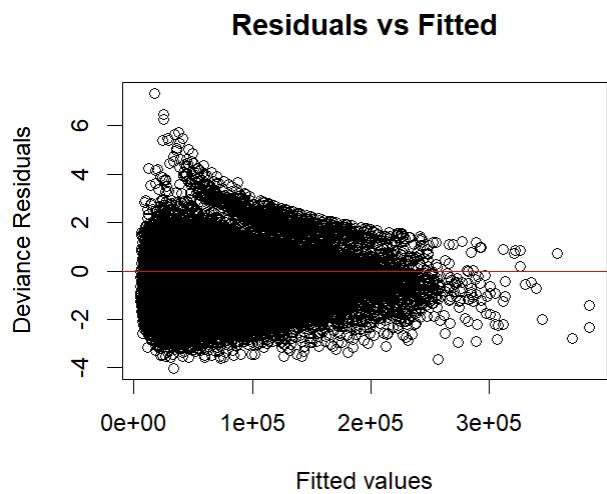
Figure 11: VIFs of the gamma log-link model

	GVIF	Df	GVIF^(1/(2*Df))
SEX	1.155623	1	1.074999
Industry	1.547901	13	1.016946
EDUC	1.382666	10	1.016333
RACE	1.426472	8	1.022449
AGE	1.085703	1	1.041971
PWSTATE2	1.420219	56	1.003137

Figure 12:

Model Fit Statistics		
Null deviance	73582	on 69999 degrees of freedom
Residual deviance	50073	on 69896 degrees of freedom
AIC	1666016	
Dispersion parameter	0.859156	

Fisher Scoring iterations	7				
Significance codes					
***	0 to 0.001				
**	0.001 to 0.01				
*	0.01 to 0.05				
.	0.05 to 0.1				
(blank)	0.1 to 1				

Variable	Estimate	Std. Error	t value	Pr(>|t|)	Significance
(Intercept)	9.2601734	0.0692016	133.814	< 2e-16	***
SEXMale	0.5399143	0.0661617	8.161	3.39e-16	***
IndustryArts, Entertainment, and Recreation, and Accommodation and Food Services	-0.3165642	0.0611031	-5.181	2.22e-07	***
IndustryConstruction	0.2536709	0.0684088	3.708	0.000209	***
IndustryEducational Services, and Health Care and Social Assistance	-0.0363058	0.0592743	-0.613	0.540205	
IndustryFinance and Insurance, and Real Estate and Rental and Leasing	0.3788696	0.0616123	6.149	7.83e-10	***
IndustryInformation	0.4023606	0.0720812	5.582	2.39e-08	***
IndustryManufacturing	0.3186779	0.0619102	5.147	2.65e-07	***

IndustryMilitary	0.2365462	0.1090978	2.168	0.030147	*
IndustryOther Services, Except Public Administration	-0.2181176	0.0633326	-3.444	0.000574	***
IndustryProfessional, Scientific, and Management, and Administrative and Waste Management Services	0.2905798	0.0606204	4.793	1.64e-06	***
IndustryPublic Administration	0.1546314	0.0629175	2.458	0.013986	*
IndustryRetail Trade	-0.1455	0.0606885	-2.397	0.01651	*
IndustryTransportation and Warehousing, and Utilities	0.1616766	0.0665703	2.429	0.015157	*
IndustryWholesale Trade	0.2741719	0.0729997	3.756	0.000173	***
EDUC1	-0.2221441	0.0768971	-2.889	0.003868	**
EDUC2	-0.1687282	0.0465954	-3.621	0.000294	***
EDUC3	-0.1960184	0.051842	-3.781	0.000156	***
EDUC4	-0.4371834	0.0455907	-9.589	< 2e-16	***
EDUC5	-0.4573674	0.041589	-10.997	< 2e-16	***
EDUC6	0.0802603	0.031742	2.529	0.011457	*
EDUC7	0.1889369	0.0326831	5.781	7.46e-09	***

EDUC8	0.3287 787	0.0333 467	9.859	< 2e-16	***
EDUC10	0.6995 114	0.0321 613	21.75	< 2e-16	***
EDUC11	1.0371 614	0.0326 706	31.74 6	< 2e-16	***
RACE2	-0.172 1167	0.0133 413	-12.9 01	< 2e-16	***
RACE3	-0.043 1591	0.0341 945	-1.26 2	0.2068 94	
RACE4	0.0535 948	0.0281 361	1.905	0.0568 05	.
RACE5	-0.176 2278	0.0748 904	-2.35 3	0.0186 18	*
RACE6	-0.027 857	0.0163 098	-1.70 8	0.0876 43	.
RACE7	-0.118 7011	0.0157 438	-7.54	4.77e- 14	***
RACE8	-0.079 9281	0.0120 993	-6.60 6	3.98e- 11	***
RACE9	-0.061 3188	0.0444 952	-1.37 8	0.1681 77	
AGE	0.0138 419	0.0002 414	57.33 3	< 2e-16	***
PWSTATE21	0.4265 536	0.0353 477	12.06 7	< 2e-16	***
PWSTATE22	0.7479 511	0.0817 063	9.154	< 2e-16	***
PWSTATE24	0.4647 916	0.0305 832	15.19 8	< 2e-16	***

PWSTATE25	0.4103 219	0.0438 273	9.362	< 2e-16	***
PWSTATE26	0.6419 116	0.0205 144	31.29 1	< 2e-16	***
PWSTATE28	0.5391 326	0.0307 616	17.52 6	< 2e-16	***
PWSTATE29	0.5221 387	0.0383 829	13.60 3	< 2e-16	***
PWSTATE210	0.4304 028	0.0698 325	6.163	7.16e-10	***
PWSTATE211	0.7812 863	0.0550 478	14.19 3	< 2e-16	***
PWSTATE212	0.4991 759	0.0226 91	21.99 9	< 2e-16	***
PWSTATE213	0.4742 87	0.0261 368	18.14 6	< 2e-16	***
PWSTATE215	0.5112 238	0.0579 586	8.82	< 2e-16	***
PWSTATE216	0.4141 822	0.0526 061	7.873	3.50e-15	***
PWSTATE217	0.4864 449	0.0249 544	19.49 3	< 2e-16	***
PWSTATE218	0.4161 129	0.0306 353	13.58 3	< 2e-16	***
PWSTATE219	0.3566 338	0.0401 467	8.883	< 2e-16	***
PWSTATE220	0.3593 998	0.0403 142	8.915	< 2e-16	***
PWSTATE221	0.3927 305	0.0360 887	10.88 2	< 2e-16	***

PWSTATE222	0.529522	0.0371027	14.272	< 2e-16	***
PWSTATE223	0.3848849	0.0578343	6.655	2.85e-11	***
PWSTATE224	0.5336249	0.0311551	17.128	< 2e-16	***
PWSTATE225	0.6104367	0.0289849	21.061	< 2e-16	***
PWSTATE226	0.39501	0.0276199	14.302	< 2e-16	***
PWSTATE227	0.4552222	0.0320691	14.195	< 2e-16	***
PWSTATE228	0.4496645	0.044499	10.105	< 2e-16	***
PWSTATE229	0.3564262	0.0307801	11.58	< 2e-16	***
PWSTATE230	0.3755509	0.0660769	5.684	1.32e-08	***
PWSTATE231	0.3708765	0.0479733	7.731	1.08e-14	***
PWSTATE232	0.6021853	0.0412501	14.598	< 2e-16	***
PWSTATE233	0.4321094	0.0557467	7.751	9.22e-15	***
PWSTATE234	0.5859758	0.0278396	21.048	< 2e-16	***
PWSTATE235	0.4357608	0.0526406	8.278	< 2e-16	***
PWSTATE236	0.5749097	0.0224667	25.589	< 2e-16	***

PWSTATE237	0.4250 627	0.0265 87	15.98 8	< 2e-16	***
PWSTATE238	0.3919 674	0.0711 357	5.51	3.60e-08	***
PWSTATE239	0.4064 094	0.0254 964	15.94	< 2e-16	***
PWSTATE240	0.3872 281	0.0389 943	9.93	< 2e-16	***
PWSTATE241	0.5486 062	0.0362 693	15.12 6	< 2e-16	***
PWSTATE242	0.4087 215	0.0251 386	16.25 9	< 2e-16	***
PWSTATE244	0.4545 78	0.0663 368	6.853	7.31e-12	***
PWSTATE245	0.3157 218	0.0336 749	9.376	< 2e-16	***
PWSTATE246	0.5479 317	0.0686 707	7.979	1.50e-15	***
PWSTATE247	0.4284 048	0.0305 171	14.03 8	< 2e-16	***
PWSTATE248	0.4798 569	0.0211 356	22.70 4	< 2e-16	***
PWSTATE249	0.4639 733	0.0390 464	11.88 3	< 2e-16	***
PWSTATE250	0.3786 196	0.0808 595	4.682	2.84e-06	***
PWSTATE251	0.5295 852	0.0275 593	19.21 6	< 2e-16	***
PWSTATE253	0.6035 55	0.0292 179	20.65 7	< 2e-16	***

PWSTATE254	0.2323105	0.0567608	4.093	4.27e-05	***
PWSTATE255	0.4236454	0.0317657	13.337	< 2e-16	***
PWSTATE256	0.4597223	0.0844814	5.442	5.30e-08	***
PWSTATE272	0.7657639	0.9273879	0.826	0.408965	
PWSTATE281	0.3916552	0.3509886	1.116	0.264485	
PWSTATE283	0.6567356	0.4640351	1.415	0.156993	
PWSTATE284	-0.7108498	0.5355324	-1.327	0.184391	
PWSTATE285	0.5746126	0.3508872	1.638	0.101511	
PWSTATE286	0.8547897	0.378913	2.256	0.024088	*
SEXMale:IndustryArts, Entertainment, and Recreation, and Accommodation and Food Services	-0.2714991	0.0705162	-3.85	0.000118	***
SEXMale:IndustryConstruction	-0.2073065	0.0763162	-2.716	0.006601	**
SEXMale:IndustryEducational Services, and Health Care and Social Assistance	-0.2543274	0.0680869	-3.735	0.000188	***
SEXMale:IndustryFinance and Insurance, and Real Estate and Rental and Leasing	-0.1419677	0.0716625	-1.981	0.047588	*
SEXMale:IndustryInformation	-0.2353771	0.0848504	-2.774	0.005538	**

SEXMale:IndustryManufacturing	-0.2814855	0.0702655	-4.006	6.18e-05	***
SEXMale:IndustryMilitary	-0.4152295	0.1193198	-3.48	0.000502	***
SEXMale:IndustryOther Services, Except Public Administration	-0.0980744	0.0742091	-1.322	0.186307	
SEXMale:IndustryProfessional, Scientific, and Management, and Administrative and Waste Management Services	-0.2406185	0.0691902	-3.478	0.000506	***
SEXMale:IndustryPublic Administration	-0.2423312	0.0730907	-3.319	0.000903	***
SEXMale:IndustryRetail Trade	-0.1511925	0.0696468	-2.171	0.029946	*
SEXMale:IndustryTransportation and Warehousing, and Utilities	-0.2027265	0.0757202	-2.677	0.007423	**
SEXMale:IndustryWholesale Trade	-0.2318317	0.0842724	-2.751	0.005943	**

Bibliography and Credits

https://www.pewresearch.org/short-reads/2025/03/04/gender-pay-gap-in-us-has-narrowed-slightly-over-2-decades/

https://news.darden.virginia.edu/2024/04/04/why-the-gender-pay-gap-persists-in-american-businesses/