Forecasting the Gender Wage Gap

Maria Casto, Manade Sok Hean, Joanna Rashid, Isabel Salazar

Instructor: Dr. Sanchita Mukherjee

ECON 138

May 20, 2021

San José State University

Abstract: In the US progress towards closing the gender wage gap lags behind other modern economies. The most significant determinants of the gender wage gap for the purpose of this paper are the educational attainment of women, marriage, rates, birth rates, and median female earnings. Regression analysis is used to forecast future values of the ratio of female to male earnings.

Table of Contents

Introduction

Unequal pay between women and men is one of the most important symbols of inequality. The gender wage gap in the United States is persisting even though women have significantly increased their earnings through their education expansion and longer working hours. The Equal Pay Act was introduced in 1963 by President John F. Kenndy. This legislation prohibits paying women lower rates for the same job strictly on the basis of their gender. However, a majority of women have still experienced discrimination in the workforce.
It is reported that women earn on average 22% less than men despite having the same education and working experience according to EPI (Wilson, 2020). However, the number is significantly larger when it comes to women of color. However, the purpose of this project is to forecast the future ratio of all women's earnings to all men's.

The main objective of this written report is to forecast the gender wage ratio in the United States. To decompose this regression analysis, we use several data points. Those data points include women's educational attainment, marriage rate, birth rate, and median earnings. This study report also presents any correlation between each variable to the gender wage gap. In addition, it includes various forecasting models with detailed explanations along with data description and model estimation. The regression analysis performed corrects for autocorrelation, heteroskedasticity, residuals, and stationarity. Additionally, some suggestions for further research are provided. These suggestions focus on economic policies which will help to minimize the wage gap in the United States.

Literature Review

Blau and Khan (Blau & Kahn, 2016) presented a study report which examined the change in the gender wage gap since the 1970s. This study explained the significant reasons behind the gender wage gap based on Blau and Khan's new empirical findings. Despite seeing a great change in the gender wage gap, especially during the technological revolution, the gender wage ratio had yet to reach 80%. Blau and Khan reported that the core success of this change resulted from the revolution of women's education and work experience up until 2010. Unfortunately, this rate of positive change could not be sustained in the long term. The forward-looking evidence in this report delineates the primary factors behind the slow change of the gender wage gap in the United States. Those factors include a family division of labor, different occupations and industries, education attainment, and so on. However, Blau and Khan also mentioned that traditional causes such as education and working experience are still salient.

The "explained" and the "unexplained" are the two main categories of gender wage within the wage decomposition in the US (Meara et al., 2019). This study explores the different factors that affect wage, which are gender segregation, unionization, parenthood, and part-time working. According to the authors, the gender wage gap tends to increase when gender segregation is being included in the estimation. Meara finds that the gap is wider among full-time workers, older workers, and unionized workers.

From 1989 to 2005, the gender wage gap in the United States was significantly narrowed and the difference between that period was about 6.4% (Suh, 2006). This was a result of diminishing the discrimination toward women in the workforce. This report specifically focuses on the factors of this change. Twelve different data points were used in this study, and four of them are quantitative variables. Therefore, several decomposition methods are being conducted simultaneously. Jingyo mentioned that the increase in working hours, education, and experience within the women population helped to narrow the wage gap. It is also noteworthy to mention that marriage status also played an important role in narrowing the wage gap in the later years. The labor force participation rate in women has also gradually increased whereas it has decreased in men. The author also concludes from his findings that gender is definitely one of the critical reasons for wage determination, but the most important improvement on this wage gap is the discrimination toward women.

Jane Waldfogel used data from the National Longitudinal Survey of Young Women in 1968-1988. She investigated the pay differences between mothers and non-mothers. The data was adjusted for part-time employment for mothers. It looked at variables such as mean wage, work experience, and potential work experience. Some alternative theories explored were Human Capital theory (lost work experience from childbearing and childrearing). According to Jane Waldfogel in The American Sociological Review in 1997 the effect of each child of an American woman has correlation to a decrease in earnings. Essentially, the top earners are generally men followed by non-mothers, and then mothers. The final reports of this study concluded that workers earned less mostly depending on their marital status along with the number of children.

According to Claudia Goldin, a Harvard Economics professor, the increasing labor force participation rate of women is helping to close the gender wage gap. Compared to 1920, the labor force participation rate for women has increased from about 55% to 75% and is climbing. Therefore, we hypothesize this will be an important variable in our model. The change in labor force participation is likely due to the increase in opportunities for women in the workforce.

A study by The World Economic Forum (2021) provides an instructive look at what must be implemented to close the gender wage gap. The forecast of the gender wage gap in this report helped to assess the accuracy of the forecast made in this paper. The World Economic Forum initially projected it would take about 100 years to close the gender wage gap. However, a recent revision to this estimate projects the gender wage gap to be eliminated in 135 years..

Description of the model

This model seeks to explain the variation in the female to male wage ratio in the United States. A review of the literature informed the choice of explanatory variables. The initial model sought out by this project is as follows:

$$fm\,ratio = child + median\,earning + marriage\,rate + percent\,bachelors + lfp\,rate + \varepsilon$$

It is expected that *child* (the average number of children born per woman) and *marriage rate* (the percent of adult women married in the US) will be negatively correlated with *fm ratio* (the ratio of female-to-male earnings). *Median earnings* (the median earnings of women in the US), *percent bachelors* (the percent of US women over age 25 with a bachelor's degree or higher), and *lfp rate* (the US labor force participation rate of women) will be positively correlated with *fm ratio*. In other words, it is expected that the more children women have and the higher their rates of marriage, the lower their earnings will be relative to men. Also, higher median earnings, higher educational attainment, and higher labor force participation of women will increase the ratio of women's wages to men’s.

It is acknowledged that median earnings and educational attainment may have endogeneity. The number of children women have on average could also be endogenous to marriage rates. However, the purpose of this model is to create a forecast of the future value of the female to male earnings ratio. For this reason, the possible endogeneity of these variables is tolerated.

Data

We collected data from three main sources: The US Census, The National Center for Education Statistics, and St. Louis Fed (FRED); all collected in a yearly format. From these sources the data dated back to 1940's, however, not all the required data for our model was available consecutively. The data set was reduced to 1993-20176 to eliminate missing values.

Data for the gender wage gap, (women) median earnings, and marriage rate were collected from the US Census. The marriage rate was not readily available from the census, therefore the data had to be transformed from available data. The marriage rate was transformed by dividing the number of women reported to be married by the number of women total for each individual year.

Data for the labor force participation rate and fertility rates were collected from the FRED. Labor force participation data was available from 1993 and later, while the fertility rate was available from as early as 1960.

Data for the percentage of women over 25 with a bachelor's degree or higher was collected from the National Center for Education Statistics. Due to the lack of data on education attainment availability for the United States, the percent of women with a bachelor's degree or higher was used instead of the breakdown of educational attainment or the median education attainment of women for given years. The data available for percent of women with a bachelor's degree or higher dated back to 1940. This data was recorded on a 10-year basis until 1990; beginning in 1995 it became available on a yearly basis.

Data Description

Once data collection was complete, each variable was plotted against time in order to visualize individual trends (Figure 2). When comparing the variables to the wage gap, a positive correlation between education and median earnings of women can be seen. The individually graphed variables also illustrate a negative correlation between marriage rate and children to the median of women's earnings.

Figure 3 confirmed the correlations that were originally assumed. The correlation matrix demonstrated a high positive correlation between female-to-male ratio and median earnings (0.901) and percent of bachelor’s degree or higher (0.866). A negative correlation between the female-to-male ratio and marriage rate (-0.898) and number of children (-0.430), confirming that as women have less children or postpone having children, their median earnings increase. The correlation between the two variables implies that the slower marriage rate of women is related to women having less children due to (0.779). Also of note in Figure 3 is the marriage rate of women and percent of women who achieve a bachelor’s degree have an inverse relationship with marriage rates and birth rates. These correlations guide the hypothesis for the model presented in this paper.

Figure 1 graphs the slowly increasing female-to-male ratio versus time. From Table 1, currently, the wage ratio is 0.805, while the starting point 28 years ago (1993) was 0.7140 noting a .09 increase in almost 30 years. It is also notable that the number of women achieving a bachelor’s degree or higher has increased by 180% from 19.2% to 34.63% in the United States from 1993 to 2017.

Model Estimation

An initial regression of the average number of children born per woman, median female earnings, percent of women that are married, female educational attainment, and female labor force participation rate on the female to male wage ratio, produce a somewhat well-fitting model. However, the highest R-squared value of 0.8934 is achieved by omitting labor force participation as an explanatory variable (Table 2). This produces the following model:

$$\underset{}{fm\,ratio} = \underset{(0.419)}{0.1709} - \underset{(0.037)}{0.009}\,child - \underset{(0.019)}{0.036}\,median\ earning + \underset{(0.643)}{0.6432}\,marriage\ rate + \underset{(0.301)}{0.6013}\,percent\ bach$$

An ADF test of this model reveals autocorrelation. For this reason, each variable is transformed with a lag value equal to one, as is prescribed by the ADF test (Table 3). The lag variables are used to create differenced variables representing the change in the variable over a period equal to one year (Table 4). The transformed values are used to create the final model.

$$fm\,ratio = \underset{(0.012)}{0.752} - \underset{(0.292)}{0.636}\,\Delta child - \underset{(0.073)}{0.078}\,\Delta median\ earning + \underset{(1.629)}{1.820}\,\Delta marriage\ rate + \underset{(0.981)}{1.676}\,\Delta percent\ ba$$

The residuals from this first-differenced model are normally distributed and indicate no autocorrelation of errors (Figure 4).

The findings of our first-differenced linear regression model indicate the average number of children per woman (*child)* has a negative impact on the female to male wage ratio. The coefficient of this variable has a high t-statistic of 2.14 and a p-value of 0.04 indicating that this value is statistically significant at a 99% level of significance. A high coefficient for *percent* *bachelors* indicates that education is very positively correlated with the ratio of female to male earnings. While the magnitude of the coefficient (1.676) is high, the t-statistic is less than 1 and the p-value is very high, indicating this coefficient is not statistically significant. The other explanatory variables in the model, *marriage rate* and *median earnings* have much smaller coefficients. They also have low t-statistics and high p-values. This indicates that while they are not statistically significant, they also do not have a significant impact on the dependent variable (Table 4).

The adjusted R-squared value for this model is only 0.286. Trial and error of other models with various combinations of explanatory variables all yielded even lower adjusted R-squared values. In consideration of the high r-squared of the original linear model, the first-differenced model discussed above was selected for use in forecasting future wage ratio values. A plot of the fitted values produced by the first-differenced model shows that they follow the same curve as the actual values closely (Figure 5, Table 5).

Results

The forecast horizon selected by this model is 10 years. This data is not seasonal and the autocorrelation of errors was eliminated by transforming the initial model into a first-differenced model. That being the case, the generic *forecast* function available in R was employed to forecast this data. The *forecast* algorithm selected ETS(A,A,N) as a forecasting method which utilizes Holt's linear method with additive errors. The sigma value for this model is 0.0169.

The forecast predicts, at a 95% confidence level, that the female to male wage ratio will be between 0.782 and 0.797 in the year 2031 (Figure 6). Using the same model the forecasted value for the year 2051 is between 0.808 and 0.822 with a 95% level of certainty.

Conclusion

The major findings of our forecast cast a dark shadow on the next generation of women and mothers joining the workforce. Even adjusting for variables such as industry, occupation, education, and experience, gender alone still accounts for roughly 51% of the difference between men's and women's earnings. Our prediction of increasing the female to male wage ratio to 82% by 2051 is indeed bleak. Our forecast only factors marital status, number of children, median earnings, and education as determinants of gender wage parity. However, it is possible that other measurable variables play a role.

An alternative theory to this conclusion would be that this is caused by something exogenous to our model. The impact of societal pressures and cultural norms that might contribute to gender wage disparities are difficult to measure. Perhaps an observational study where participants answer questions that ask how these pressures influence their choices on variables like number of children, level of education, and marital status. Further research on factors affecting the gender wage gap might seek to quantify these yet unexplained variables.

Our findings are not readily transferable to other countries. The US lag behind most other industrialized economies in terms of social services and welfare support for women and families. However, a comparison of the gender wage gap in the US to other modern economies would serve to identify policies that effectively increase gender wage parity.

Ultimately, our findings indicate that progress toward gender wage parity in the US has largely stalled. In order to avoid dooming the next generation of women to the status quo, the US must continue to research its unique obstacles to gender equality and adopt radical policy change.

Tables and Figures

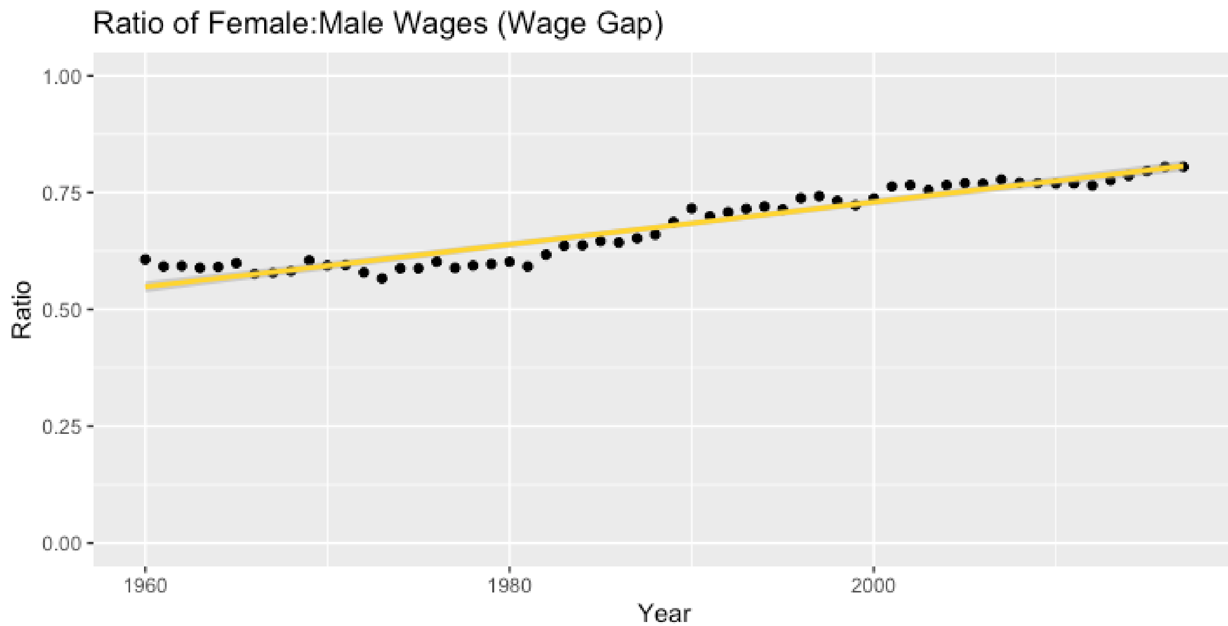

Figure 1. The ratio of female-to-male wage gap from 1960 to 2017 with a trendline.

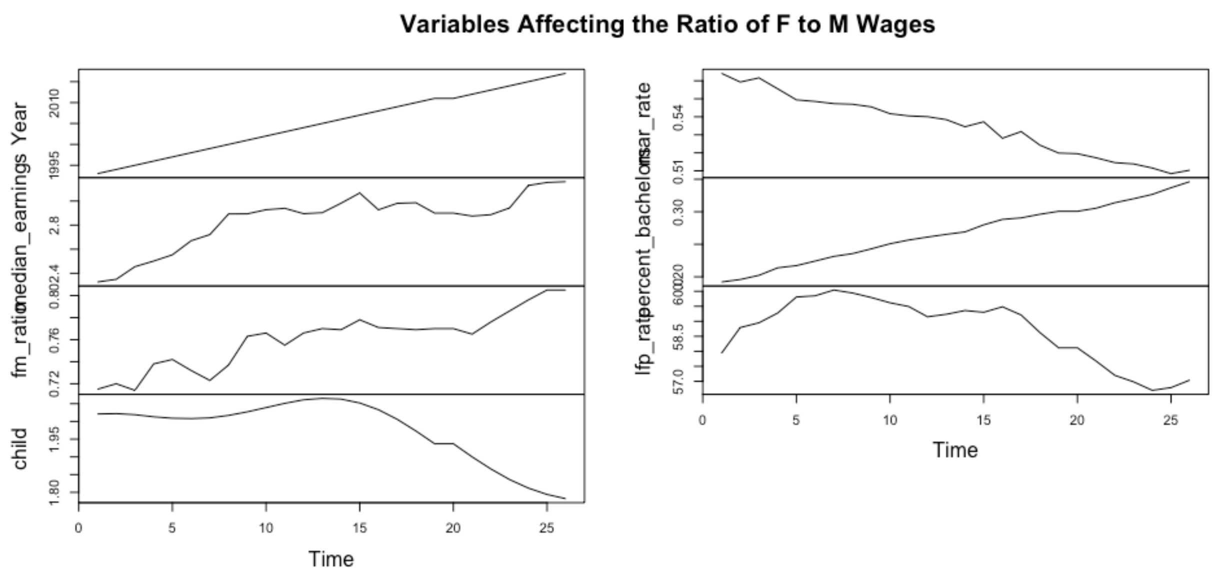

Figure 2. Female- to-male wage gap model variables graphed over time.

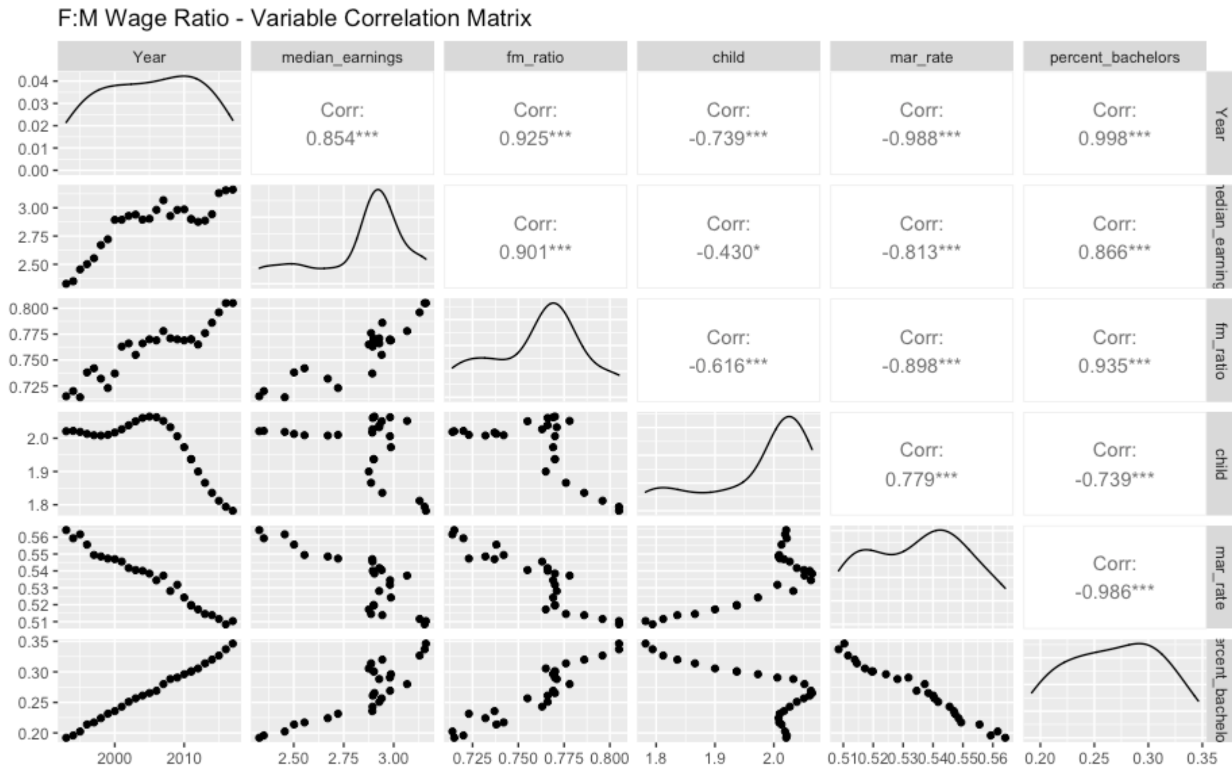

Figure 3. Variable correlation matrix for the wage gap.

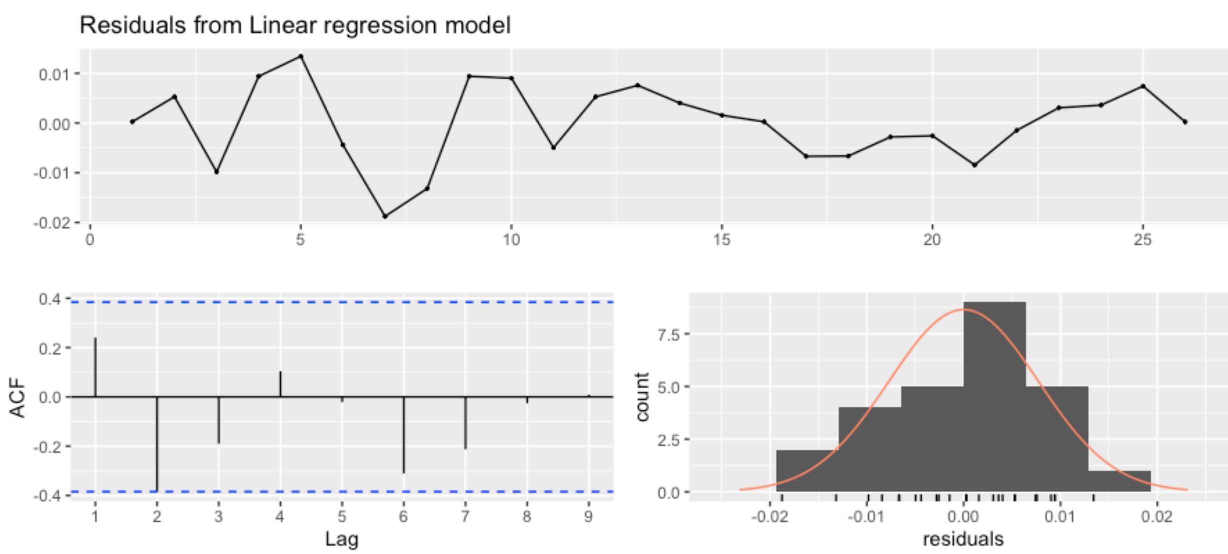

Figure 4. Residuals from linear regression.

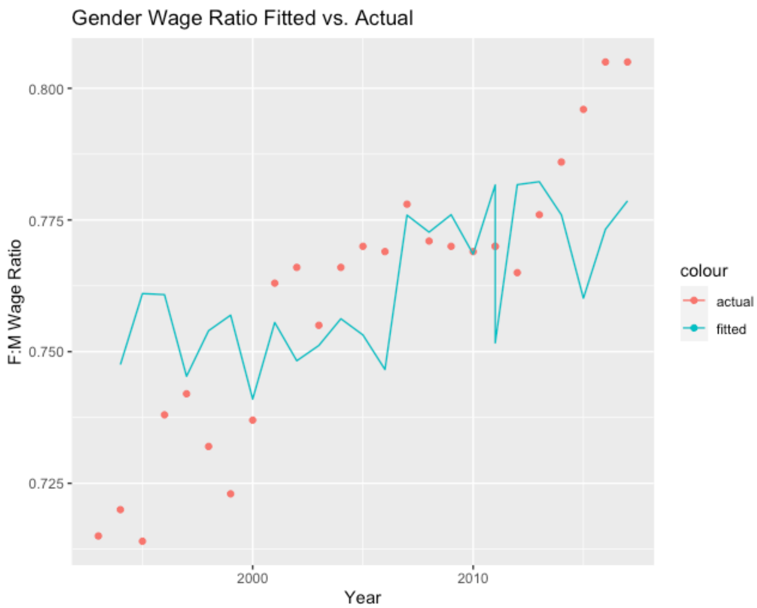

Figure 5. Actual versus fitted gender wage gap.

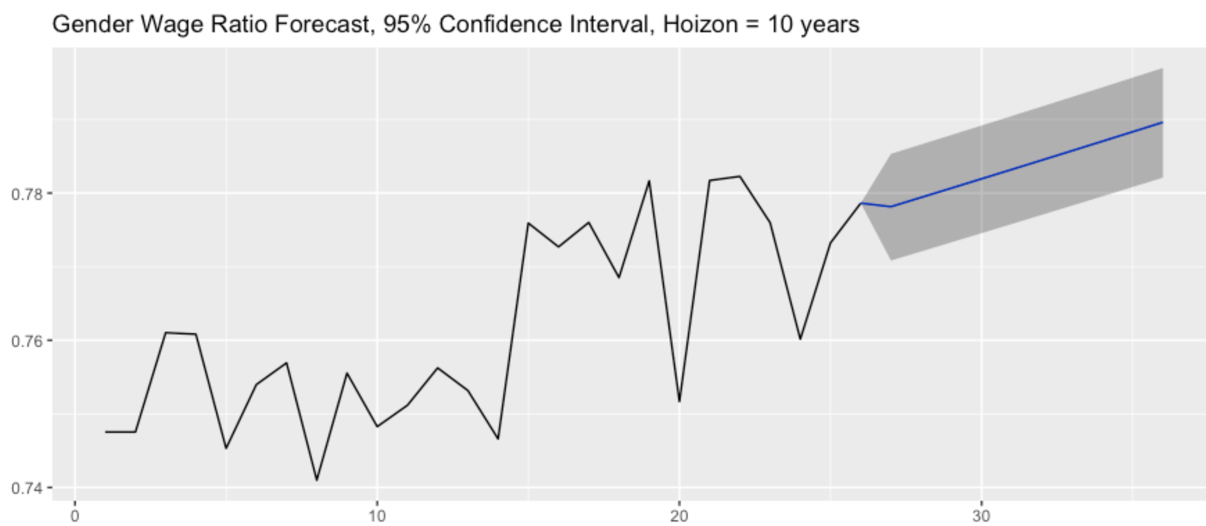

Figure 6. Gender wage gap forecast, with 95% confidence level for a 10-year horizon.

Table 1. **Explanatory Variable Summary Statistics**

Female Median Earnings	Female:Male Wage Ratio	Average number of children per woman	Marriage Rate	Percent Women achieving bachelor's degree or higher	Labor Force Participation Rate
Min. :2.328	Min. :0.7140	Min. :1.782	Min. :0.5084	Min. :0.1920	Min. :56.70
1st Qu.:2.760	1st Qu.:0.7390	1st Qu.:1.937	1st Qu.:0.5196	1st Qu.:0.2325	1st Qu.:57.98
Median :2.900	Median :0.7675	Median :2.011	Median :0.5378	Median :0.2672	Median :59.18
Mean :2.843	Mean :0.7604	Mean :1.975	Mean :0.5350	Mean :0.2679	Mean :58.71
3rd Qu.:2.972	3rd Qu.:0.7708	3rd Qu.:2.031	3rd Qu.:0.5472	3rd Qu.:0.3007	3rd Qu.:59.49
Max. :3.161	Max. :0.8050	Max. :2.065	Max. :0.5641	Max. :0.3463	Max. :60.04

Table 2. Initial Linear Regression Model Summary

Residuals:				
Min	1Q	Median	3Q	Max
-0.0187647	-0.0048	0.0003	0.0053	0.0134
term	estimate	std.error	statistic	p.value
(Intercept)	0.1709	0.4192	0.4077	0.6876
child	-0.0088	0.0370	-0.2369	0.8150
mar_rate	0.6432	0.6948	0.9258	0.3651
median_earnings	0.0357	0.0194	1.8403	0.0799
percent_bachelors	0.6013	0.3007	1.9995	0.0587
Residual standard error:	0.008425 on 21 degrees of freedom			
Multiple R-squared:	0.9104	Adjusted R-squared:	0.8934	
F-statistic:	53.36 on 4 and 21 DF	p-value:	0.0000	

Table 3. ADF Test for Autocorrelation

term	Dickey-Fuller Value	Lag Order	p.value
fm_ratio	-2.9529	1	0.2094
child	-1.1905	1	0.8808
mar_rate	-2.304	1	0.4566
median_earnings	-2.258	1	0.474
percent_bachelors	-2.6719	1	0.3164
alternative hypothesis:	stationary		

Table 4: First-Differenced Model Linear Regression Model Summary:

Residuals:				
Min	1Q	Median	3Q	Max
-0.047029	-0.0117	0.0005	0.0168	0.0359
term	estimate	std.error	statistic	p.value
(Intercept)***	0.752338172	0.01223156433	61.5079275	2.86E-24
child*	-0.6364689534	0.2916051556	-2.182639577	0.0411568785
mar_rate	1.820669923	1.629678108	1.117196037	0.2771522545
median_earnings	-0.07824632106	0.07293385844	-1.072839457	0.2961180944
percent_bachelors	1.676128592	1.707824531	0.9814407521	0.3380995848
Residual standard error:	0.02274 on 20 degrees of freedom			
Multiple R-squared:	0.2867	Adjusted R-squared:	0.1440	
F-statistic:	2.009 on 4 and 20 DF	p-value:	0.1319	

Table 5. Fitted vs actual values

Fitted Versus Actual					
Year	fm_ratio	fitted values	Year	fm_ratio	fitted values
1993	0.715	NA	2006	0.769	0.747
1994	0.720	0.748	2007	0.778	0.776
1995	0.714	0.761	2008	0.771	0.773
1996	0.738	0.761	2009	0.770	0.776
1997	0.742	0.745	2010	0.769	0.769
1998	0.732	0.754	2011	0.770	0.782
1999	0.723	0.757	2011	0.770	0.752
2000	0.737	0.741	2012	0.765	0.782
2001	0.763	0.756	2013	0.776	0.782
2002	0.766	0.748	2014	0.786	0.776
2003	0.755	0.751	2015	0.796	0.760
2004	0.766	0.756	2016	0.805	0.773
2005	0.770	0.753	2017	0.805	0.779

References

Blau, F. D., & Kahn, L. M. (2016, January). *The gender wage gap: extent, trends, and explanations*. nber.org. https://www.nber.org/system/files/working_papers/w21913/w21913.pdf.

Census Bureau. (2020, December). *MS-1. Marital status of the population 15 years old and over by sex, race and hispanic origin: 1950 to present*. [Data set]. United States Census Bureau. https://www.census.gov/data/tables/time-series/demo/families/marital.html

Goldin, C., Related Content, Horwitz, S., Summers, L. H., Small, K. A., & Jeffrey M. Jones and Thomas MaCurdy. (n.d.). Gender Gap. Retrieved from https://www.econlib.org/library/Enc/GenderGap.html

Fontenot, K., Semega, J., & Kollar, M. (2018, September 12). *Table-A-4. Number and real median earnings of total workers and full-time, year-round workers by sex and female-to-male earnings ratio: 1960 to 2017*. [Data set]. United States Census Bureau. https://www.census.gov/library/publications/2018/demo/p60-263.html

FRED, Federal Reserve Bank of St. Louis. (2021, April 27). *Fertility rate, total for the United States (SPDYNTFRTINUSA)*. World Bank. https://fred.stlouisfed.org/series/SPDYNTFRTINUSA

FRED, Federal Reserve Bank of St. Louis. (2021, May 7). *Labor force participation rate - women (LNS11300002)*. [Data set]. United States Bureau of Labor Statistics. https://fred.stlouisfed.org/series/LNS11300002#0

Meara, K., Pastore, F., & Webster, A. (2019, September 5). *The gender pay gap in the USA: a matching study*. Journal of Population Economics. https://link.springer.com/article/10.1007/s00148-019-00743-8.

National Center for Education Statistics. (2019, October). *Table 104.10. Rates of high school completion and bachelor;s degree attainment among persons age 25 and over, by race/ethnicity and sex: Selected years 1910 through 2019*. [Data set]. United States Census Bureau & United States Department of Commerce. https://nces.ed.gov/programs/digest/d19/tables/dt19_104.10.asp

Suh, J. (2006). *Decomposition of the change in the gender wage gap*. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.495.6697&rep=rep1&type=pdf

Waldfogel, J. (1997, April). The effect of children on women's wages'. *American Sociological Association, 62*(2), 209-217. https://doi.org/10.2307/2657300

Wilson, V. (2020, March 30). *Exposed and underpaid: Women still make less than men, including in sectors especially affected by the coronavirus*. Economic Policy Institute. https://www.epi.org/blog/exposed-and-underpaid-women-still-make-less-than-men-including-in-sectors-especially-affected-by-the-coronavirus/.

World Economic Forum. (2021, March 30). *Global gender gap report 2021*. World Economic Forum. http://www3.weforum.org/docs/WEF_GGGR_2021.pdf