3/11/2022

ECON 550 Final Paper

An Analysis on the Effect of Having Children on the Gender Wage Gap

Chaudhari, Kasrovi, Reichlin

Introduction

There is an abundance of economic literature surrounding trends in the gender wage gap in the United States. Empirical evidence suggests that since the 1950's, labor force participation rates of women have substantially increased. Labor force participation is not the only area in which women are seeing increased levels of representation. Data drawn from the Current Population Survey's Annual Social and Economic Supplement show an increasing gender gap in both enrollment and graduation rates in college (Parker, Pew Research Center). Women today are more likely to receive bachelor's degrees than men. Yet, evidence of the gender wage gap still exists. In *The Gender Wage Gap: Extent, Trends, and Explanations,* Blau and Kahn show that the disparity in men's and women's wages has shrunk over time but remains higher at the top level of earnings (Blau & Kahn, 2017). In an attempt to gain a better insight into this wage disparity among men and women, we will study the effects of having a child on wages and how this effect differs between men and women. I.e., what is the effect of childbearing on the gender wage gap?

A considerable amount of research on the gender wage gap exists, as well as on the effect of marriage and childbearing on an individual's earnings. A study analyzing panel data beginning in 1979 suggests that early marriage and childbearing can lead to a decrease in women's wages, while marriage and children have little to no effect on men's wages (Loughran & Zissimopoulos, 2009). Traditional gender roles may offer an explanation for these findings. Historically, women have been more likely to stay home and assume the role of primary childcare giver. The expectation of pregnancy and maternity leave, materializing as a reduction in labor productivity, is one logical explanation of the observed gender wage gap in past years. More recently, traditional gender roles in American homes have become less common. Women are more likely to serve as their household's primary breadwinner than they have been in the past. Our analysis will differ from the mentioned literature in that we will use cross-sectional data drawn from a sample taken in 2019. Using a sample from 2019 provides us with relatively current data that are not subject to the exogenous employment shocks resulting from COVID-19.

Our assumption, due to the presence of the gender wage gap and the potential temporary reduction in labor productivity resulting from pregnancy, is that childbearing is associated with a lower expected wage for women than for men, holding productivity and demographic variables fixed. Determining the extent to which childbearing affects the gender wage gap can be a useful tool in identifying the proportion of wage disparities among men and women that can be explained away through the aforementioned reduction in labor productivity. In the following section of this paper, we specify the models used in our analysis, along with our key assumptions and economic theory used to derive our model specifications. In subsequent sections, we provide an overview of the data used in our analysis, followed by a discussion of our findings, a brief conclusion, and potential methods to further our research.

Data Summary

Prior to analyzing our model specifications, a brief summary of the data can serve as a useful tool for subsequent analysis. The data used in our analysis is drawn from IPUMS's American Community Survey (ACS). A brief summary of variables can be found in Table 1.

Table 1: Summary of Variables

Variable Name	**Description**
incwage	Annual pre-tax wage and salary income
educ_simple	Level of educational attainment
age	Age of respondent
married	If respondent is married (dummy variable)
female	If respondent is female (dummy variable)
metro	If respondent lives in a metropolitan area (dummy variable)
classwkrd	Class of worker (e.g., self-employed)
occ	Occupation
uhrswork	Usual hours worked per week
child_pres	If respondent has a child
chlt5_pres	If respondent has a child under 5 years old
nchild	Number of children in household

The dependent variable in our analysis, representing earned income, required little data manipulation past accounting for missing values and individuals with no annual earned income. The independent variables of primary interest to our model include *female* and *married*. *Marst,* originally coded as a categorical variable detailing marital status, has been renamed and re-coded as a dummy variable. The re-coded variable, *married*, takes on a value of ‘1’ if the respondent is married, and ‘0’ if the respondent is single, separated, or divorced. *Female,* originally *sex,* has been recoded to take on a value of ‘1’ if the respondent is female, and ‘0’ if the repsondent is male. Coding these variables in this way will allow us to capture the differential effect in wages of married women relative to married men in subsequent regression analyses.

To correct for omitted variable bias in the coefficients on our variables of interest, we include a set of control variables aimed to capture demographic and productivity characteristics. *Educd,* renamed as *educ_simple* has been re-coded in such a way as to report the respondent’s level of educational attainment. The educational attainment categories include no highschool, some highschool, highschool diploma or equivalent, some college, Associate’s Degree, Bachelor’s Degree, Master’s Degree, Professional Degree, and Doctoral Degree. Treating education as a categorical variable rather than a continuous variable detailing the respondent’s years of education avoids any potential bias that may otherwise be introduced into the model through misspecification of educational attainment. *Age* required virtually no data manipulation past limiting our sample to individuals considered to be prime age, or 25 to 54 years old. In subjecting our sample to this constraint, we hope to eliminate individuals who may be working low-wage jobs while in school, as well as older individuals who may be retiring in the coming years. While *age* and *educ_simple* are far from perfect measures of a respondent’s level of productivity, their inclusion in the model likely serve as necessary control variables.

The *occ* variables provide detailed occupational categories on over 500 distinct occupations. Categories including unemployed and military specific occupations have been recoded as missing values.

Nchild required no data manipulation, although we did use it to create a new binary variable to indicate if the respondent has at least one child. Similarly, *chlt5_pres* required some data manipulation in recoding it as a binary indicator of whether or not the respondent has a child less than five years old present.

Our primary geographic variable, *metro*, originally included five distinct classifications: "Metropolitan status indeterminable (mixed)," "Not in metropolitan area," "In metropolitan area - In central/principal city," "In metropolitan area - Not in central/principal city," and "In metropolitan area - Central/principal city status indeterminable (mixed)." We have re-coded *metro* into a dummy variable, with a value of '0' indicating the respondent does not live in or near a metropolitan area, and a value of '1' indicating that they do live in or near a metropolitan area. The ACS uses county, or other available geographic information, to determine which metropolitan classification to assign to each respondent. Because of this, we have decided to categorize "Metropolitan status indeterminable (mixed)" as the respondent living in or near a metropolitan area. Tables 2 through 7 provide summary statistics on our included variables.

Table 2: Variable Summary Statistics

Variable	**No. Obs.**	**Mean**	**Median**	**SD**	**Min.**	**Max.**
incwage	922,704	60708.64	45000	67535.37	4	717000
age	922,704	39.4633	39	8.731813	25	54
uhrswork	922,704	40.76113	40	11.253	1	99
nchild	922,704	1.035268	1	1.204479	0	9

Table 3: Variable Features (Dummy)

Variable	**Marital Status**	**No. Obs.**
married	0 (single)	41.41%
	1 (married)	58.59%
Total		**922,704**

Table 4: Variable Features (Dummy)		
Variable	**Gender**	**No. Obs.**
female	0 (male)	51.73%
	1 (female)	48.27%
Total		**922,704**

Table 5: Variable Features (Dummy)		
Variable	**If Respondent Lives in Metropolitan Area**	**No. Obs.**
metro	0 (Not in/near metropolitan area)	8.86%
	1 (In/near metropolitan area)	91.14%
Total		**922,704**

Table 6: Variable Features (Dummy)		
Variable	**If Respondent Has A Child**	**%**
child_pres	0 (No children)	46.71%
	1 (At least 1 child)	53.29 %
Total		**922,704**

Table 7: Variable Features (Dummy)		
Variable	**If Respondent Has A Child Under 5**	**%**
chlt5_pres	0 (No children under 5)	83.19%
	1 (At least 1 child under 5)	16.81%
Total		**922,704**

Model Specifications

To analyze the effect having children has on the gender wage gap, we initially regress individuals' annual wages on a number of control variables and our independent variables of interest. Our first specification takes the following form:

$$
\begin{aligned}
incwage_i = \beta_0 + \beta_1 married_{\boldsymbol{i}} + \beta_2 female_{\boldsymbol{i}} + \beta_3 child\_pres_{\boldsymbol{i}} + \beta_4 (child\_pres_i * female_{\boldsymbol{i}}) \\
+ \beta_5 age_i + \beta_6 age_{\boldsymbol{i}}^{\boldsymbol{2}} + \beta_7 metro_{\boldsymbol{i}} + \beta_8 uhrswork_{\boldsymbol{i}} + \beta_9 occ_{\boldsymbol{i}} + \beta_{10} classwkrd_{\boldsymbol{i}} \\
+ \beta_{11} educ\_simple_i + \varepsilon_i
\end{aligned}
$$

β_4 is our coefficient of interest, representing the wage differential of women with at least one child relative to men with at least one child. β_2 represents the proportion of the gender wage gap that is independent of whether or not an individual has a child, while β_3 effectively captures the effect of having a child on men's wages. The remaining regressors included in the model serve as controls to mitigate omitted variable bias in the coefficient of interest. *Age,* serving in part as a crude proxy for years of work experience, has a direct positive influence on wages. Additionally, there is a positive relationship between one's age and whether or not they have a child. These two conditions indicate that the variable's omission would cause the coefficient of interest to be biased in the positive direction. We have chosen to implement a non-linear specification of *age,* as years of experience in the workforce likely has a marginal effect on income that begins to diminish at a certain point, before hitting a threshold that is associated with a reduction in wages for each additional year of age. Further discussion on functional form specification is provided in later sections of this paper.

Marital status, having a direct positive influence on wages, as well as a positive relationship with *child_pres,* has also been included to control for the associated bias that would be present if omitted. The regressor indicating whether or not the respondent lives in, or near to, a metropolitan area is included in the model to control for omitted variable bias. We expect those who live in cities or suburban areas to, on average, have a higher wage than those who live in rural areas. Our data show that there is a weakly negative relationship between the presence of at least one child and metropolitan city status. These two conditions would materialize in a biased coefficient on *child_pres* in the negative direction. Similarly, the coefficient on the variable representing usual hours worked per week has been included to mitigate a negative bias that would otherwise confound our estimate of the causal effect of children on wages. It is plausible that the more hours a person works, the higher their annual earned income will be. We have also found there to be a negative relationship between hours worked per week and the likelihood of having a child present.

In addition to the control variables previously discussed, we include occupational category, class of worker, and highest level of educational attainment to control for their respective fixed effects. Accounting for occupational fixed effects controls for any variation in earnings among occupations. Certain occupations like computer and mathematical jobs, or construction tend to have a greater male representation. Controlling for occupational categories also captures any risk aversion in job selection that may differ among men and women. This would otherwise cause a difference in samples for different occupational categories and hence can have different measures for the income in those categories

While we do expect that the presence of a child is associated with a negative wage differential for women relative to men, it is plausible that women's wage patterns eventually recover from the reduction in labor productivity stemming from childbirth. Our second specification, in which we have replaced *child_pres* with a dummy variable indicating whether or not the respondent has a child less than five years old, takes the following form:

$$\begin{aligned} incwage_i = \beta_0 + \beta_1 married_i + \beta_2 female_i \ + \beta_3 chlt5\_pres_i + \ \beta_4 (chlt5\_pres_i * female_i) \\ + \ \beta_5 nchild_i + \beta_6 age_i + \beta_7 age_i^2 + \beta_8 metro_i + \beta_9 uhrswork_i + \beta_{10} occ_i \\ + \beta_{11} classwkrd_i + \beta_{12} educ\_simple_i \ + \varepsilon_i \end{aligned}$$

This model shares many of the characteristics described previously, with two key distinctions. As mentioned, our independent variable of interest now indicates whether the respondent has a child less than five years old. A comparison of coefficients on child indicator variables, provided in a subsequent section, will not permit the information necessary to test the assumption that women's wage patterns return to normal five years post-childbirth, however a comparison of the magnitude of the coefficients on each model's interaction term can provide us with information on the proportion of the gender wage gap that is attributable to childbearing across both models. The second difference comes from the inclusion of a variable representing the number of children a respondent has. This variable is not included in our initial model due to collinearity restraints, but warrants inclusion to effectively control for any prior labor productivity reductions resulting from former pregnancies.

Analysis

Prior to providing analysis on the regression outputs for our two models, we will briefly discuss our choice of functional form specification. To test whether our model omits any polynomial terms, we conduct a RESET Test by estimating a model with no polynomial terms to collect predicted values of income. We run a subsequent regression with two additional inclusions, a variable representing our squared predicted values and a variable representing our cubed predicted values, followed by a joint hypothesis test. The results of this test are displayed below:

```
HDFE Linear regression                            Number of obs   =    922,704
Absorbing 3 HDFE groups                           F(   9, 922155) =   11770.72
                                                  Prob > F        =     0.0000
                                                  R-squared       =     0.3766
                                                  Adj R-squared   =     0.3762
                                                  Within R-sq.    =     0.1006
                                                  Root MSE        = 53338.3406
```

incwage	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
age	-586.7076	35.98137	-16.31	0.000	-657.2299	-516.1854
metro	-4325.764	300.6008	-14.39	0.000	-4914.932	-3736.597
uhrswork	-623.1153	38.83283	-16.05	0.000	-699.2263	-547.0043
married	-3277.263	224.8163	-14.58	0.000	-3717.895	-2836.631
female						
Female	5446.945	379.1593	14.37	0.000	4703.806	6190.085
child_pres						
At least 1 child	-10205.32	553.9966	-18.42	0.000	-11291.13	-9119.504
female#child_pres						
Female#At least 1 child	9907.987	532.0258	18.62	0.000	8865.234	10950.74
c.yhat#c.yhat	.0000238	7.13e-07	33.42	0.000	.0000224	.0000252
c.yhat#c.yhat#c.yhat	-1.03e-10	4.08e-12	-25.30	0.000	-1.11e-10	-9.51e-11
_cons	48618.6	1619.953	30.01	0.000	45443.55	51793.66

H_0: $\boldsymbol{\delta}_1 = 0$, $\boldsymbol{\delta}_2 = 0$

```
test _b[c.yhat#c.yhat] = _b[c.yhat#c.yhat#c.yhat] = 0

( 1)  c.yhat#c.yhat - c.yhat#c.yhat#c.yhat = 0
( 2)  c.yhat#c.yhat = 0
      Constraint 2 dropped

      F(  1,922155) = 1116.85
           Prob > F =    0.0000
```

The F-statistic and corresponding p-value on the joint hypothesis test that the coefficients on our quadratic and cubic variables are equal to zero indicate that the null hypothesis should be rejected at the 99% significance level. I.e., we reject the null hypothesis that the model with no polynomial terms is specified correctly.

As touched on briefly in the previous section, we choose a non-linear specification of *age* based on the theory that marginal labor productivity increases diminish at a certain age before hitting a threshold that is associated with a marginal decrease in labor productivity. It is important to note that RESET Tests do not indicate which variables are misspecified. We considered a non-linear specification of *uhrsworked*, as it is reasonable to assume that there are diminishing returns to productivity after a certain number of hours worked per week. A quadratic, or even a cubic, specification could be justified in this case. We have decided against a non-linear specification for usual hours worked per week on the basis that for individuals in jobs that pay hourly wages, *uhrswork* will have a linear marginal effect on income. It is also possible that the correct specification of *age* takes on a higher order polynomial than included in our model. A further discussion of potential functional form misspecification is included at the end of this section.

Model 1

We model our regression using robust standard error estimation to account for heteroskedasticity**.** Let us consider the case where we don't impose any sample restrictions. Holding all else equal, the presence of a child is associated with an estimated increase in income of $12,186 for men and an increase of $625 for women. The wage differential for women having a child present is -$11,561, relative to men with a child present. The coefficient on *female* represents the estimated gender wage gap that is independent of whether or not the respondent has a child present.

$$incwage_i = \beta_0 + \beta_1 married_i + \beta_2 female_i + \beta_3 child\_pres_i + \beta_4 (child\_pres_i * female_i) + \beta_5 age_i + \beta_6 age_i^2 + \beta_7 metro_i + \beta_8 uhrswork_i + \beta_9 occ_i + \beta_{10} classwkrd_i + \beta_{11} educ\_simple_i + \varepsilon_i$$

$$\frac{\partial incwage}{\partial child\_pres} = \beta_3 + \beta_4 female$$

```
HDFE Linear regression                          Number of obs     =    922,704
Absorbing 3 HDFE groups                         F(   8, 922156)   =    9151.56
                                                Prob > F          =     0.0000
                                                R-squared         =     0.3748
                                                Adj R-squared     =     0.3745
                                                Within R-sq.      =     0.0980
                                                Root MSE          = 53414.3167
```

incwage	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
married	4969.994	111.9004	44.41	0.000	4750.673	5189.315
female						
Female	-8790.07	164.5421	-53.42	0.000	-9112.567	-8467.573
child_pres						
At least 1 child	12186.39	193.1515	63.09	0.000	11807.82	12564.96
female#child_pres						
Female#At least 1 child	-11561.1	225.0741	-51.37	0.000	-12002.23	-11119.96
age	2299.281	64.82431	35.47	0.000	2172.228	2426.335
c.age#c.age	-17.41507	.8398182	-20.74	0.000	-19.06109	-15.76906
metro	7112.153	134.3425	52.94	0.000	6848.846	7375.46
uhrswork	1095.942	6.686302	163.91	0.000	1082.837	1109.047
_cons	-54713.21	1225.496	-44.65	0.000	-57115.14	-52311.28

In the second case, we impose the sample restriction on incwage. We think that including the extreme outliers in the data of incwage can give us biased estimates. In this case we therefore create a new variable that only contains data for incwage between its 1st and 99th percentiles. In this case, holding all else equal, the effect of the presence of a child on income is an increase by $8510 for men and an decrease by $177 for women. The wage differential for women having a child present is -$8,687, relative to men with a child present. We can also see that for this case the R-squared value increased from .375 to .44. This implies that a larger amount of variation in the dependent variable is explained in this case after imposing sample restrictions compared to the previous one.

Lastly, we drop more outliers that may be present to get a better range for *incwage*. In this case we create a new variable that only contains data for incwage between its 5th and 95th percentiles. In this case, holding all else equal, the effect of the presence of a child on income is an increase of $5,240 for men and a decrease of $1,020 for women. The wage differential for women having a child present is -$6,260, relative to men with a child present. We can also see that for this case the R-squared has improved slightly compared to the previous case.

Model 2

We model our regression using robust standard errors to account for any heteroskedasticity that may be present. Let us consider the case where we don't impose any sample restrictions. Holding all else equal, the estimated effect of the presence of at least one child under five years old on income is an increase of $2,058 for men and an increase of $1,246 for women. We can see that the increase is less compared to that of the men. The wage differential for women is -$812, relative to men both having at least one child under 5.

$$incwage_i = \beta_0 + \beta_1 married_i + \beta_2 female_i + \beta_3 chlt5\_pres_i + \beta_4(chlt5\_pres_i * female_i) + \beta_5 nchild_i + \beta_6 age_i + \beta_7 age_i^2 + \beta_8 metro_i + \beta_9 uhrswork_i + \beta_{10} occ_i + \beta_{11} classwkrd_i + \beta_{12} educ\_simple_i + \varepsilon_i$$

$$\frac{\partial incwage}{\partial chl5t\_pres} = \beta_3 + \beta_4 female$$

```
HDFE Linear regression                          Number of obs   =    922,704
Absorbing 3 HDFE groups                         F(   9, 922155) =    8115.45
                                                Prob > F        =     0.0000
                                                R-squared       =     0.3730
                                                Adj R-squared   =     0.3726
                                                Within R-sq.    =     0.0954
                                                Root MSE        = 53492.0181
```

incwage	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
married	5871.444	111.3352	52.74	0.000	5653.231	6089.657
female						
Female	-14204.73	154.263	-92.08	0.000	-14507.08	-13902.38
chlt5_pres						
At least 1 child under 5	2058.065	276.897	7.43	0.000	1515.356	2600.773
female#chlt5_pres						
Female#At least 1 child under 5	-811.6935	309.8778	-2.62	0.009	-1419.044	-204.3434
nchild	2276.148	60.10704	37.87	0.000	2158.341	2393.956
age	2187.867	65.38987	33.46	0.000	2059.705	2316.029
c.age#c.age	-15.40734	.8446854	-18.24	0.000	-17.06289	-13.75178
metro	7182.202	134.5059	53.40	0.000	6918.575	7445.829
uhrswork	1118.639	6.704836	166.84	0.000	1105.498	1131.78
_cons	-51826.99	1240.827	-41.77	0.000	-54258.97	-49395.01

In the second case, we impose the sample restriction on incwage. In this case we create a new variable that only contains data for incwage between its 1st and 99th percentiles. In this case, holding all else equal, the effect of the presence of at least one child under 5 on income is an increase of $2,373 for men and an increase of $1,297 for women. We can see that the increase for women is less compared to that of men. The wage differential for women having a child less than five years old is -$1,076, relative to men. We can also see that for this case the R-squared value increased from .37 to .44. This implies that a larger amount of variation in the dependent variable is explained in this case after imposing sample restrictions compared to the previous one.

Lastly, we drop more outliers that may be present to get a better range for incwage. In this case we create a new variable that only contains data for incwage between its 5th and 95th percentile. In this case, the effect of the presence of at least one child under five on income is an increase of $2,251 for men and an increase by $816 for women. The increase in wages is less for women compared to men. The wage differential for women is -$1,435, relative to men both having at least one child under 5. We can also see that for this case the R-squared has improved slightly compared to the previous case.

Table 8: Regression Result Comparisons

	Female Coefficient	Child Indicator Coefficient	Interaction Coefficient	R-Squared
Model 1 (No Restrictions)	-8,790.07 (164.5421)	12,186.39 (193.1515)	-11,561.1 (225.0741)	0.3748
Model 1 (1^{st} & 99^{th} Pct. Dropped)	-6,248.703 (118.088)	8,510.552 (128.5101)	-8,687.174 (154.8617)	0.4417
Model 1 (5^{th} & 95^{th} Pct. Dropped)	-3,809.47 (83.746)	5,239.303 (86.3176)	-6,260.061 (107.8542)	0.4553
Model 2 (No Restrictions)	-14,204.73 (154.263)	2,058.065 (276.897)	-811.6935 (309.8778)	0.373
Model 2 (1^{st} & 99^{th} Pct. Dropped)	-10,230.57 (105.027)	2,373.215 (176.2826)	-1,076.12 (207.5335)	0.4395
Model 2 (5^{th} & 95^{th} Pct. Dropped)	-6,565.225 (71.1407)	2,251.759 (114.58)	-1,435.148 (142.6164)	0.453

Note: Figures in parentheses below coefficient estimates indicate heteroskedasticity robust standard errors. All coefficient estimates included in this table are statistically significant at the 99% significance level.

Table 9: Summary of Incwage With Restrictions

	No. Obs.	Mean	Median	SD	Min.	Max.
1^{st} & 99^{th} Pct. Dropped	902,358	56,655.47	45,000	48,139.05	1,100	420,000
5^{th} & 95^{th} Pct. Dropped	833,008	51,690.12	45,000	32,428.33	6,000	154,000

Limitations

There are a number of limitations that pose potential threats to the internal validity of our analysis. As previously mentioned, experimenting with higher order polynomial specifications of *age* and *uhrswork* may change our key estimates. If the intuition surrounding our specification decisions does not hold in the real world, it could result in biased coefficient estimates. An additional potential source of concern stems from possible measurement error in the data. It is important to consider this possibility with self-reported income measures. It could be the case that respondents systematically under or over report their earnings. However we do not have a reason to believe that this type of error in our dependent variable is correlated with any of our included regressors. If this assumption holds true, measurement error in our data will simply result in estimated standard errors that are larger than they would otherwise be. It will not cause our slope coefficient estimates to be biased and will not impede drawing causal inferences from the regression output.

A more concerning threat comes from the possibility of simultaneity causality bias being present in the model. If in reality, the decision to have children is a function of individuals' income, it would necessarily cause bias in our estimated coefficients. A potential remedy would be to include an instrumental variable in our model, although we were unable to select a variable that is both strongly relevant and only influences the dependent variable through our child indicator variables. If these two conditions do not hold, the inclusion of an instrumental variable would likely produce a greater degree of bias than the bias present from simultaneity. The final potential source of bias relates to any variables we may have omitted in our model. A variable representing years of work experience would be an important inclusion, however the ACS does not collect data on this. It is likely that there are other variables contributing to omitted variable bias that we have looked over when specifying our models.

Conclusion

Despite the increase in women's labor force participation rates, our analysis clearly suggests that the gender wage gap remained prevalent in 2019. Our initial model, using the presence of at least one child as the child indicator variable, showed that the wage differential for women having children relative to men is associated with a larger reduction in wages than the gender proportion of the gender wage gap that is independent of childbearing. This result was not unexpected, but it does suggest that childbearing is not the sole cause of the observed wage disparity among genders. The coefficients of interest estimated in our second model were in the direction that we expected them to be, however the magnitude of the coefficient on the interaction term was a surprise. The wage differential of women with at least one child less than five years old present relative to men was considerably smaller than the estimated coefficient on the interaction term in the previous model. This result was not in line with our assumption that the wage differential in the second model would be greater in magnitude than in the first. One avenue to extend our research involves incorporating additional variables relating to children, e.g., age of children and age of mother during childbirth.

To further our research, a significant improvement in our work can be made using panel data instead of cross sectional data. This would enable us to employ a Differences-in-Differences estimation and let us test our assumption that women's wage patterns would return to normal in the years following childbirth. Additionally, we could control for any time-varying fixed effects, as well as individual level fixed effects. This can help us answer other important questions like if women's earnings eventually catch up later in their lives compared to men or if having children has a permanent effect on their earnings until they retire.

References

Blau, Francine & Kahn, Lawrence (2017) “The Gender Wage Gap: Extent, Trends, and Explanations.” *Journal of Economic Literature, 55(3), 789-865.* *https://pubs.aeaweb.org/doi/pdfplus/10.1257/jel.20160995*

Loughran, David S. & Zissimopoulos, Julie M. (2009) “Why Wait? The Effect of Marriage and Childbearing on the Wages of Men and Women.” *The Journal of Human Resources,* Vol. 44, No. 2, pp. 326-349. https://www-jstor-org.ezproxy2.library.drexel.edu/stable/20648900?seq=8#metadata_info_tab_contents

Parker, Kim. “What's behind the Growing Gap between Men and Women in College Completion?” *Pew Research Center*, Pew Research Center, 8 Nov. 2021, https://www.pewresearch.org/fact-tank/2021/11/08/whats-behind-the-growing-gap-between-men-and-women-in-college-completion/