Understanding the Effect of Age-Gap of Children on Female Income

Heewon Oh 301268860 December 6, 2021

ECON 435 Term Paper

Introduction

Significant research supports the existence of a substantial gender wage-gap (see Figure 1) and supports some of the factors that cause the differences in female wages to that of. While it is known that the expected earnings of female workers fall when they have children, no significant results have affirmed any relationship between the age-gap between the youngest and oldest siblings on the expected earnings of the mother. Thus, this paper seeks to investigate this relationship between maternal wage and hours worked with the age-gap of children. Our initial intuition is that is that the greater the age-gap, the lower the wage due to greater time out of the workforce, which results in less human capital accumulation. We believed that this may be partially offset by greater hours worked due to greater flexibility when there is an older child to help supervise a younger one.

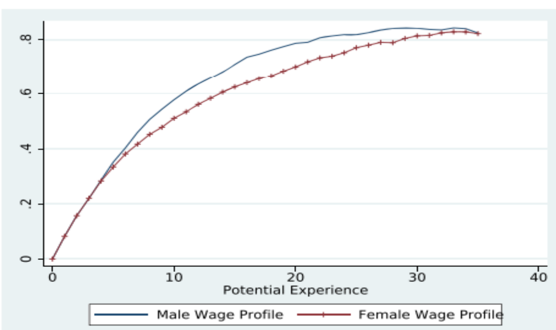

Figure 1: Log wage vs potential experience by gender (Sauer & Taber, 2021)

Furthermore, this paper builds upon a study conducted by Sauer and Taber titled "Understanding Women's Wage Growth Using Indirect Inference with Importance Sampling." We drew our inspiration for our research questions after reviewing the Markov Model results for human capital, marriage, and fertility models. Sauer and Taber's paper found that the difference in the curvature in log female wage growth to that of male's is driven primarily by the difference in human capital accumulation (Sauer & Taber, 2021). This is greatly due to married and fertile women opting to work less. The paper found that wages among prime-age women would be 2.4% higher if the relationship between fertility and employment status were removed. The wages would also be 4.7% higher if the relationship between fertility and marriage were also removed.

Background

Overall, we were unsuccessful in finding credible studies that support significance in the relationship between the age-gap of siblings on maternal hourly wage and hours worked. In addition, indirectly, previous studies suggest that this may challenging to find. Daniel et al (2013) shows that, on average, women's earnings depreciate per child. Thus, given that the number of children is undoubtedly correlated with age-gap, any results by the age-gap could have potential confounding by the number of children.

In addition, it is widely noted that women choose occupations with lower depreciation of human capital (Polachek, 1981), thus minimizing the magnitude of minor factors under the guise of the human capital model posed by Sauer and Taber.

Sauer and Taber (2021) proposed that human capital accumulation (i.e. work experience, education, etc.) is the key model that can be used to explain the wage and wage-growth. They used a Markov model, meaning that it is static.

$$\mathcal{S}_{it} \equiv \{t,\, L_{it},\, M_{it},\, H_{it},\, K_{it},\, \{A_{1it},\, ..,\, A_{K_{it}it}\};\, E_i,\, \nu_i\}$$

Figure 2: State Variables under Sauer and Taber's Markov model

$$\dot{H} = -\delta H$$

Figure 3: Depreciation of Human Capital

$$\dot{H} = a(\mathcal{S}_{it})(\bar{H}_i - H_{it})e^{-\mu_i t}$$

Figure 4: Human capital accumulation function

In Figure 2, the State Variables that effect the human capital accumulation include A1it,…Akit, which are the ages of the children. Thus, it is reasonable to inquire whether the age-difference between the oldest and youngest child influences the state variables, which in turn influences the wage rate; this in combination with social influences, may affect the number of hours worked. This is due to the equations in figure 3 and 4, where the depreciation of human capital plays a non-linear effect on human capital accumulation. However, this premises on the assumption of the depreciation effect as well as the non-linear effect relationship, or a difference in employment status depending on the age gap that is not attributable to other explained regressors.

Data

We used the unbalanced panel dataset provided by Sauer and Taber's paper for reproducibility. We used the main file containing 52 variables and over 1 million observations. Notable variables included: id number, marriage status, employment status, years of education, number of children, age, number of birth month and year of children, wage, race, wave number, and year and SIPP record.

The data was compiled using female data from unbalanced panel data from SIPP (Survey of Income and Program Participation) by the US Census Bureau from the 1996, 2001, 2004, and 2008 reports. This is a survey program that attempts to create panel data via interviewing participants as many times as possible through a four-year period via 4-month interviews.

	ed <dbl>	married <dbl>	numkids <dbl>	age <dbl>	potexperience <dbl>	agediff <dbl>	divorced <dbl>
median	13.000000	1.0000000	1.000000	34.00000	20.00000	4.000000	0.0000000
mean	13.656182	0.5528627	1.280322	33.61940	19.96322	4.901118	0.1032634
SE.mean	0.002859	0.0006309	0.001642	0.01137	0.01118	0.009154	0.0003861
CI.mean.0.95	0.005603	0.0012366	0.003219	0.02229	0.02191	0.017942	0.0007568
var	5.075617	0.2472059	1.675405	80.30562	77.62430	16.928878	0.0926002
std.dev	2.252913	0.4971981	1.294374	8.96134	8.81047	4.114472	0.3043029
coef.var	0.164974	0.8993156	1.010976	0.26655	0.44133	0.839497	2.9468600

Figure 5: Descriptive Stats Table for Key Factor Variables compiled on dataset with 621033 observations.

We attempted to recreate the filtered dataset used by Sauer and Taber's paper to have a comparable reference. Nevertheless, the descriptive statistics for several key variables varied from the results found in the paper. We believe some of the discrepancy is from filtering and some of it is from differences in age-calculation and weighting the unbalanced panel data.

Limitations

Unfortunately, we were unable to use the large ancillary data and models that Sauer and Taber created due to limitations of the scope of this paper and our resources. As such, we were unable to obtain individualized estimations of human capital levels, wage growth and fertility under the respective models.

Therefore, we were not able to directly compare maternal wages and hours worked to human capital levels and fertility. Thus, we can only compare direct effects between the age difference of children to maternal income.

Ultimately, a show of significance using the variables available would help support the argument that the age-difference of the oldest and youngest child plays a role in determining maternal income; however, non-significant results do not necessarily sufficiently support the counterargument. Moreover, in the latter case, research must be conducted to compare with key models.

Methodology

We evaluate the effect of the age-gap between the oldest and youngest child on the log wage of women and the hours worked by women using the static unbalanced panel.

Thus, we use the equations below for determining the log wage:

Logwi = B0 + educationB1 + numberofkidsB2 + potentialexperienceB3 + marriedB4+divorcedB5+hoursB6+agedifferenceB7 + ageB8 + ai + eit

And the hours worked:

Hours = B0 + educationB1 + numberofkidsB2+ potentialexperienceB3 + marriedB4+agedifferenceB5 + ai + eit

In the equations, B0 represents the intercept, potentialexperience the potential years in the labor force, agedifference the difference between the ages of the youngest and oldest child, ai the unobservable individual fixed effects and eit epsilon for time-dependent shocks.

We use the plm package in R to test using the Fixed-Effect, Random-Effect and Between Models. We then use the Hausman test to compare the random-effect test to that of the fixed-effect, assuming the fixed-effect model is accurate.

Results

Overall, we included 6 regression outputs and two Hausman Tests: one for each model for each equation and a Hausman Test to compare the Fixed-Effect and Random-Effect models.

Figures 6-9: Regression results for factors on log wage using various models and Hausman Test.

```
Twoways effects Within Model

Call:
plm(formula = logwg ~ ed + potexperience + numkids + age + married +
    divorced + hrs + agediff + as.factor(calyr), data = white,
    effect = "twoways", model = "within", index = c("id"))

Unbalanced Panel: n = 17092, T = 1-15, N = 123885

Residuals:
     Min.   1st Qu.    Median   3rd Qu.      Max.
-2.979807 -0.062198  0.000345  0.066658  3.109102

Coefficients: (3 dropped because of singularities)
                       Estimate Std. Error t-value Pr(>|t|)
potexperience         0.002678   0.002639    1.01  0.31024
married               0.008895   0.006312    1.41  0.15875
divorced              0.020526   0.007179    2.86  0.00425 **
hrs                  -0.003613   0.000135  -26.77  < 2e-16 ***
as.factor(calyr)97    0.003187   0.005573    0.57  0.56743
as.factor(calyr)98    0.034707   0.007787    4.46  8.3e-06 ***
as.factor(calyr)99    0.062715   0.009756    6.43  1.3e-10 ***
as.factor(calyr)100   0.059102   0.013053    4.53  6.0e-06 ***
as.factor(calyr)104   0.018489   0.010378    1.78  0.07481 .
as.factor(calyr)105   0.010956   0.008716    1.26  0.20877
as.factor(calyr)106   0.002085   0.007155    0.29  0.77073
as.factor(calyr)108   0.069341   0.014966    4.63  3.6e-06 ***
as.factor(calyr)109   0.050224   0.013648    3.68  0.00023 ***
as.factor(calyr)110   0.030869   0.012431    2.48  0.01302 *
as.factor(calyr)111   0.011617   0.011088    1.05  0.29479
as.factor(calyr)112   0.007932   0.009354    0.85  0.39646
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    7830
Residual Sum of Squares: 7750
R-Squared:      0.0112
Adj. R-Squared: -0.147
F-statistic: 75.5448 on 16 and 106763 DF, p-value: <2e-16
```

Figure 6

```
Oneway (individual) effect Between Model

Call:
plm(formula = logwg ~ ed + numkids + age + married + divorced +
    hrs + potexperience + agediff + as.factor(calyr), data = white,
    model = "between", index = c("id"))

Unbalanced Panel: n = 17092, T = 1-15, N = 123885
Observations used in estimation: 17092

Residuals:
   Min. 1st Qu.  Median 3rd Qu.    Max.
-2.5209 -0.2770 -0.0232  0.2753  3.5378

Coefficients: (1 dropped because of singularities)
                      Estimate Std. Error t-value Pr(>|t|)
(Intercept)           0.117641   0.033906    3.47  0.00052 ***
ed                    0.101918   0.001682   60.61  < 2e-16 ***
numkids               0.011686   0.004539    2.57  0.01005 *
age                   0.011679   0.000565   20.67  < 2e-16 ***
married               0.176084   0.012053   14.61  < 2e-16 ***
divorced              0.099496   0.015211    6.54  6.3e-11 ***
hrs                   0.010666   0.000345   30.90  < 2e-16 ***
agediff              -0.004086   0.001001   -4.08  4.5e-05 ***
as.factor(calyr)97    0.067526   0.035695    1.89  0.05854 .
as.factor(calyr)98    0.055036   0.036715    1.50  0.13388
as.factor(calyr)99    0.097588   0.034729    2.81  0.00496 **
as.factor(calyr)100  -0.043519   0.070574   -0.62  0.53748
as.factor(calyr)104   0.165267   0.020850    7.93  2.4e-15 ***
as.factor(calyr)105   0.159957   0.027995    5.71  1.1e-08 ***
as.factor(calyr)106   0.129895   0.032095    4.05  5.2e-05 ***
as.factor(calyr)107   0.059354   0.037023    1.60  0.10892
as.factor(calyr)108   0.159519   0.030063    5.31  1.1e-07 ***
as.factor(calyr)109   0.203131   0.027361    7.42  1.2e-13 ***
as.factor(calyr)110   0.145111   0.035103    4.13  3.6e-05 ***
as.factor(calyr)111   0.171574   0.041470    4.14  3.5e-05 ***
as.factor(calyr)112   0.034081   0.044081    0.77  0.43945
as.factor(calyr)113   0.157528   0.054330    2.90  0.00374 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    5210
Residual Sum of Squares: 3490
R-Squared:      0.331
Adj. R-Squared: 0.33
F-statistic: 401.7 on 21 and 17070 DF, p-value: <2e-16
```

Figure 7

```
Oneway (individual) effect Random Effect Model
   (Swamy-Arora's transformation)

Call:
plm(formula = logwg ~ ed + numkids + age + married + divorced +
    hrs + potexperience + agediff + as.factor(calyr), data = white,
    model = "random", index = c("id"))

Unbalanced Panel: n = 17092, T = 1-15, N = 123885

Effects:
                 var std.dev share
idiosyncratic 0.0726  0.2694   0.3
individual    0.1714  0.4140   0.7
theta:
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.455   0.761   0.798   0.777   0.815   0.834

Residuals:
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 -2.826  -0.095   0.013   0.005   0.116   3.052

Coefficients: (1 dropped because of singularities)
                      Estimate Std. Error z-value Pr(>|z|)
(Intercept)           0.535003   0.027600   19.38  < 2e-16 ***
ed                    0.106956   0.001610   66.42  < 2e-16 ***
numkids              -0.001125   0.004378   -0.26  0.79715
age                   0.014153   0.000514   27.56  < 2e-16 ***
married               0.044259   0.005541    7.99  1.4e-15 ***
divorced              0.033689   0.006496    5.19  2.2e-07 ***
hrs                  -0.001612   0.000127  -12.73  < 2e-16 ***
agediff              -0.003654   0.000972   -3.76  0.00017 ***
as.factor(calyr)97    0.014771   0.003701    3.99  6.6e-05 ***
as.factor(calyr)98    0.050486   0.003988   12.66  < 2e-16 ***
as.factor(calyr)99    0.083569   0.004263   19.60  < 2e-16 ***
as.factor(calyr)100   0.081285   0.007779   10.45  < 2e-16 ***
as.factor(calyr)104   0.133120   0.008490   15.68  < 2e-16 ***
as.factor(calyr)105   0.138102   0.008614   16.03  < 2e-16 ***
as.factor(calyr)106   0.129486   0.008816   14.69  < 2e-16 ***
as.factor(calyr)107   0.131174   0.009601   13.66  < 2e-16 ***
as.factor(calyr)108   0.153367   0.009508   16.13  < 2e-16 ***
as.factor(calyr)109   0.139364   0.009065   15.37  < 2e-16 ***
as.factor(calyr)110   0.132099   0.009237   14.30  < 2e-16 ***
as.factor(calyr)111   0.114577   0.009396   12.19  < 2e-16 ***
as.factor(calyr)112   0.110101   0.009569   11.51  < 2e-16 ***
as.factor(calyr)113   0.103455   0.010234   10.11  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    12200
Residual Sum of Squares: 9200
R-Squared:      0.246
Adj. R-Squared: 0.246
Chisq: 9253.21 on 21 DF, p-value: <2e-16
```

Figure 8

```
        Hausman Test

data:  hrs ~ potexperience + ed + numkids + married + agediff
chisq = 24, df = 2, p-value = 6e-06
alternative hypothesis: one model is inconsistent
```

Figure 9

Figures 10-13: Regression results for factors on hours using various models and Hausman Test.

```
Oneway (individual) effect Between Model

Call:
plm(formula = hrs ~ potexperience + ed + numkids + married +
    agediff + as.factor(calyr), data = white, model = "between",
    index = c("id"))

Unbalanced Panel: n = 18664, T = 1-15, N = 147266
Observations used in estimation: 18664

Residuals:
   Min. 1st Qu.  Median 3rd Qu.    Max.
 -38.16   -8.21    2.79    8.38   65.91

Coefficients:
                    Estimate Std. Error t-value Pr(>|t|)
(Intercept)          14.3201     0.8513   16.82  < 2e-16 ***
potexperience         0.2072     0.0149   13.87  < 2e-16 ***
ed                    0.7736     0.0417   18.54  < 2e-16 ***
numkids              -1.9341     0.1181  -16.38  < 2e-16 ***
married              -3.2632     0.2295  -14.22  < 2e-16 ***
agediff               0.0810     0.0267    3.03  0.00246 **
as.factor(calyr)97    4.5252     1.1948    3.79  0.00015 ***
as.factor(calyr)98    1.6400     1.5167    1.08  0.27957
as.factor(calyr)99    5.5535     1.4841    3.74  0.00018 ***
as.factor(calyr)100  -5.8225     3.0879   -1.89  0.05937 .
as.factor(calyr)104  11.4425     0.5892   19.42  < 2e-16 ***
as.factor(calyr)105  12.4903     0.7881   15.85  < 2e-16 ***
as.factor(calyr)106   8.5895     0.9103    9.44  < 2e-16 ***
as.factor(calyr)107  10.6118     1.0503   10.10  < 2e-16 ***
as.factor(calyr)108  11.5686     0.8392   13.79  < 2e-16 ***
as.factor(calyr)109   9.5064     0.7674   12.39  < 2e-16 ***
as.factor(calyr)110  10.9247     0.9876   11.06  < 2e-16 ***
as.factor(calyr)111  11.4900     1.1684    9.83  < 2e-16 ***
as.factor(calyr)112  11.4884     1.2464    9.22  < 2e-16 ***
as.factor(calyr)113   7.3896     1.5642    4.72  2.3e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    3610000
Residual Sum of Squares: 3020000
R-Squared:      0.164
Adj. R-Squared: 0.163
F-statistic: 192.452 on 19 and 18644 DF, p-value: <2e-16
```

Figure 10

```
Twoways effects Random Effect Model
   (Swamy-Arora's transformation)

Call:
plm(formula = hrs ~ potexperience + ed + numkids + married +
    agediff, data = white, effect = "twoways", model = "random",
    index = c("id"))

Unbalanced Panel: n = 18664, T = 1-15, N = 147266

Effects:
                   var  std.dev share
idiosyncratic 6.32e+01 7.95e+00  0.27
individual    1.67e+02 1.29e+01  0.73
time          8.37e-03 9.15e-02  0.00
theta:
         Min. 1st Qu. Median    Mean 3rd Qu.   Max.
id    0.47609  0.7875 0.8177 0.7960  0.8252 0.8432
time  0.07151  0.3646 0.3996 0.3747  0.4216 0.4354
total 0.06037  0.3227 0.3857 0.3662  0.4109 0.4315

Residuals:
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  -40.9    -8.9     5.5     0.1    10.0    77.5

Coefficients:
               Estimate Std. Error z-value Pr(>|z|)
(Intercept)    16.36068    0.09573   170.9   <2e-16 ***
potexperience   0.27235    0.00166   163.7   <2e-16 ***
ed              1.02696    0.00540   190.1   <2e-16 ***
numkids        -2.21180    0.01546  -143.0   <2e-16 ***
married        -1.99927    0.01597  -125.2   <2e-16 ***
agediff         0.12949    0.00354    36.6   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    36600000
Residual Sum of Squares: 34100000
R-Squared:      0.0693
Adj. R-Squared: 0.0693
Chisq: 111681 on 5 DF, p-value: <2e-16
```

Figure 11

```
Twoways effects Within Model

Call:
plm(formula = hrs ~ potexperience + ed + numkids + married +
    agediff + smagediff + lgagediff + midagediff + as.factor(calyr),
    data = white, effect = "twoways", model = "within",
    index = c("id"))

Unbalanced Panel: n = 18664, T = 1-15, N = 147266

Residuals:
   Min. 1st Qu.  Median 3rd Qu.    Max.
 -62.76   -1.82    0.00    1.97   79.07

Coefficients: (2 dropped because of singularities)
                     Estimate Std. Error t-value Pr(>|t|)
potexperience          0.2704     0.0711    3.80  0.00014 ***
married               -1.2502     0.1494   -8.37  < 2e-16 ***
as.factor(calyr)97    -0.0661     0.1422   -0.46  0.64234
as.factor(calyr)98     0.2616     0.2025    1.29  0.19643
as.factor(calyr)99     0.9185     0.2565    3.58  0.00034 ***
as.factor(calyr)100    1.6387     0.3389    4.84  1.3e-06 ***
as.factor(calyr)104   -0.2611     0.2846   -0.92  0.35879
as.factor(calyr)105   -0.4719     0.2403   -1.96  0.04957 *
as.factor(calyr)106   -0.4422     0.2005   -2.21  0.02745 *
as.factor(calyr)108    0.2798     0.4188    0.67  0.50409
as.factor(calyr)109   -0.1366     0.3826   -0.36  0.72104
as.factor(calyr)110   -0.4996     0.3514   -1.42  0.15505
as.factor(calyr)111   -0.3578     0.3181   -1.12  0.26063
as.factor(calyr)112   -0.0256     0.2717   -0.09  0.92508
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    8140000
Residual Sum of Squares: 8120000
R-Squared:      0.00156
Adj. R-Squared: -0.144
F-statistic: 14.388 on 14 and 128574 DF, p-value: <2e-16
```

Figure 12

```
        Hausman Test

data:  logwg ~ ed + numkids + age + married + divorced + hrs + potexperience +  ...
chisq = 2439, df = 16, p-value <2e-16
alternative hypothesis: one model is inconsistent
```

Figure 13

Moreover, we see that for both equations, that the between and random effects model shows significance for the age-difference variable while the fixed-effect model does not. Also, as we can see, the equations are independent of the effects of time variables. We also see that the adjusted R-squared for the within model is largely very small while it is reasonably significant on others. The F-stat values are above 10 each all regressions. We also note that critically, in the between and random effects models, the magnitude for the age-difference regressor is negative in hourly wage and positive in hours worked. This is consistent with our initial intuition.

Furthermore, we see that for both equations, the Hausman test shows that the fixed and random-effect models are certainly inconsistent with each other. Thus, the regressions are certainly puzzling at initial glance as they deliver an ambivalent result for significance.

Conclusion

Unfortunately, due to the limitations of our models and our results, we have inconclusive results surrounding the effects of the age-gap of children on the mother's wage and hours worked.

We received polarizing results between the fixed-effect compared to the between and random-effect models, even after fixing factors such as time. This is not entirely surprising given that we used an unbalanced panel dataset. Moreover, the fixed-effect and between methods give less weighting on individuals with fewer T. Given this, we ran the Hausman test, which supported the alternative hypothesis, despite efforts to address endogeneity. By standard, we used the fixed-effects results and thus conclude that we were unable to find any significance between age-gap and both hours worked and hourly wage of mother.

Nevertheless, this is not to say that our results support the conclusion that age-gap does not play a determinant factor in maternal income. This is because given the models developed by Sauer and Taber, human capital and fertility are key to determining women's wage and wage growth. It is possible that there exists a significant correlation between age-gap to these models. Intuitively, it seems probable given the design of the models. Unfortunately, due to the limitations of this paper, we are unable to draw on estimations of these values. Ultimately, while we are unable to use our random effects results, we anticipate that we could draw more conclusive results by comparing to values drawn from the human capital and fertility models.

References

Sauer, R. M., Taber, C.. Understanding women's wage growth using indirect inference with importance sampling. J Appl Econ. 2021;36:453–473. https://doi.org/10.1002/jae.2818

Sauer, R. M., Taber, C.(2021). Supplemental Appendix for Understanding Women's Wage Growth using Indirect Inference with Importance Sampling

Oesch, D., Lipps, O., & McDonald, P. (2017). The wage penalty for motherhood: Evidence on discrimination from panel data and a survey experiment for Switzerland. *Demographic Research*, *37*, 1793-1824. https://doi.org/10.4054/demres.2017.37.56

Polachek, S. W. (1981). Occupational self-selection: A human capital approach to sex differences in occupational structure. *The Review of Economics and Statistics*, *63*(1), 60. https://doi.org/10.2307/1924218