Mother Wage Penalty

Across industrialized countries, it is a well-studied phenomenon that childless women are paid more on average than mothers. In this exercise, we use survey data to investigate how the structural aspects of jobs affect the wages of mothers relative to the wages of childless women.

The exercise is based on: Wei-hsin Yu and J anet Chen-Lan Kuo. 2017. The Motherhood Wage Penalty by Work Conditions: How Do Occupational Characteristics Hinder or Empower Mothers?" *American Sociological Review* 82(4): 744-769.

In this paper, the authors examine the association between the so-called *mother wage penalty* (i.e., the pay gap between mothers and non-mothers) and occupational characteristics. Three prominent explanations for the motherhood wage penalty–"stressing work-family conflict and job performance," "compensating differentials," and "employer discrimination"–provide hypotheses about the relationship between penalty changes and occupational characteristics. The authors use data from 16 waves of the National Longitudinal Survey of Youth to estimate the effects of five occupational characteristics on the mother wage penalty and to test these hypotheses.

This paper uses a type of data known as 'panel data.' Panel data consist of observations on the same people over time. In this example, we are going to analyze the same women over multiple years. When analyzing panel data, each time period is referred to as a *wave*, so here each year is a wave. The most general form of model for working with panel data is the *two-way fixed effects model*, in which there is a fixed effect each woman and for each wave.

The data file is `yu2017sample.csv`, which is a CSV file. The names and descriptions of variables are:

Name	Description
PUBID	ID of woman
year	Year of observation
wage	Hourly wage, in cents
numChildren	Number of children that the woman has (in this wave)
age	Age in years
region	Name of region (North East = 1, North Central = 2, South = 3, West = 4)
urban	Geographical classification (urban = 1, otherwise = 0)
marstat	Marital status
educ	Level of education
school	School enrollment (enrolled = TRUE, otherwise = FALSE)
experience	Experience since 14 years old, in days
tenure	Current job tenure, in years
tenure2	Current job tenure in years, squared
fullTime	Employment status (employed full-time = TRUE, otherwise = FALSE)
firmSize	Size of the firm
multipleLocations	Multiple locations indicator (firm with multiple locations = 1, otherwise = 0)
unionized	Job unionization status (job is unionized = 1, otherwise = 1)
industry	Job's industry type
hazardous	Hazard measure for the job (between 1 and 2)
regularity	Regularity measure for the job (between 1 and 5)
competitiveness	Competitiveness measure for the job (between 1 and 5)
autonomy	Autonomy measure for the job (between 1 and 5)
teamwork	Teamwork requirements measure for the job (between 1 and 5)

Question 1

What years are included in the data? How many women are included, and how many person-years are included? Briefly comment on the results.

Question 2

Create a new variable called `logwage` that is the log of `wage`. Explain why it might be important to use such a variable in the study at hand. Calculate and comment on the correlations between `logwage`, `numChildren`, `experience`, and at least two other continuous variables in the data, e.g., `competitiveness`, `autonomy`, `teamwork`.

Question 3

Run a two-way fixed effects effects (factor(region) + factor(year)) regression where the outcome is the log wage and the predictor is the number of children that a woman has. Include fixed effects for each *woman* and each *year* (i.e., include factorized variables for `year` and `PUBID` in the regression model). What is the estimated coefficient on the variable `numChildren`? Provide a brief substantive interpretation of the coefficient.

Question 4

What is the standard error of the coefficient for `numChildren`? What is the value of the estimate divided by the standard error and what does that mean? If you have the null hypothesis that this coefficient is equal to zero, at what level of α would you still be able to reject the null hypothesis? What is the meaning of rejecting the null hypothesis in this exercise? Based upon Question 3, are there other predictors we might want to include in this model? For what reasons might we want to include them?

Question 5

Keeping in fixed effects for *woman* and *year*, introduce the following variables in the regression model:

- Location (`region`, `urban`)
- Marital Status (`marstat`)
- Human Capital (`educ`, `school`, `experience`, `tenure`, `tenure2`)
- Job Characteristics (`fullTime`, `firmSize`, `multipleLocations`, `unionized`, `industry`)

Report the coefficient and standard error of `numChildren` now. Is the coefficient statistically significant? Provide a brief substantive interpretation of this coefficient and the coefficients for any two other variables.

Question 6

In order to replicate the original study, add interactions between `numChildren` and `regularity` and between `numChildren` and `hazardous`. Report the five coefficients involving these variables and their standard errors. Is it statistically significant? Interpret the interaction term for `numChldren` and `hazardous`. Based on this model, what change would you expect to see in the log wages of two otherwise identical mothers, if the first had one more child than the second? Can we interpret the effect of occupation characteristics on motherhood wage penalty as causal? Why or why not?