OptOut project: studying maternal employment sequences

This file explains the documents that constructed the dataset and produced the analysis.

**Sample Selection - Python**

Code file:

* 1sample\_selection.ipynb

Output file:

* analytic\_df.csv – main model sample, mothers who gave first births before <= 1995
* analytic\_df\_seq14.csv – 14-year-seq sample, larger than main model sample, include mothers who gave first births before <= 1999

Content outline

* Drop males
* Drop discontinued subsamples
* Use dob reported in 1979
* The following categories are dropped, but do not count as missing. We can’t observe them anyways
  + Drop women who remained childless and remained in the study until the end of their reproductive years (age 35)
  + Drop women who were childless and died before age 35
  + Drop women who had children and died before child turned 18
* The following categories are dropped, but count as missing
  + Women who had first birth before 1995, but quit study before child turned 18
* Get final sample size: 3434
* Repeat the steps and get 14-year-seq sample size 3682

**Monthly Employment Sequences Construction - Python**

Code file: 2seq\_construction.ipynb

Output files:

* full\_wk\_df.csv: transformed weekly work statuses, from week 1 to week 1879
* monthly\_df.csv: monthly statuses for 14 years after first childbirth

Content outline

Decisions about missing data

* **Within-job gaps that started or ended prior to 1/1/1978**: three such cases, involving three respondents. Disregard these gaps – see detailed reasons in notebook
* **Within-job gaps with missing start/end date (missing other than -4, valid skip)**: 225 unique respondents have some gaps with missing start/end dates. We have a total of 302 cases with missing start/end dates, so some respondents have multiple missing dates. Among the 302 cases, 139 had only partial missing – only start is missing, or only end is missing, and only 12 of them had a start/end date within 4 months of a birth. So we decided we only include within-job gaps when we have both valid start and end dates. The share of gaps with missing dates among all gaps is 2%. The share of respondents with missing gaps among all respondents is 7%.
* **Respondent 8511, with a missing birth month**: her missing birth month is no problem at all.

Transform weekly statuses

* Identify weeks of maternity leaves from within-job gaps. Any gap that overlaps with the birth month or started within 16 weeks (4 months) after the last week of birth month is considered maternity leave. We do not consider any gaps prior to 1/1/1978 (only three cases) or any gaps with missing start/end date. No need to worry about unidentified leaves for moms who gave birth before 1978
* Assign working, with work hour 0 to part-time (previously we assigned them to working with missing hours)
* Use weekly labor force status and working hours to determine other statuses:

|  |  |  |  |
| --- | --- | --- | --- |
|  | Customized Status | NLSY Status | NLSY Hours |
| 0 | missing | 0 | 0 |
| 1 | full-time employed | {Job number 100-2510, 3 – assoc. with emp, gap dates missing, all time not acctd for} | >= 35 |
| 7 – active military service | 0 |
| 2 | working with missing hours | {Job number 100-2510, 3} | <= 0 |
| 3 | part-time employed | {Job number 100-2510, 3} | [0, 35) |
| 4 | nonworking | {2, 4, 5} | <= 0 and part-time  Respondent with id 1951 have 5 weeks of part-time hour, but work status {4, 5} |
| 5 | maternity leave | Identified as prev step | |

* Construct monthly sequences from weekly sequences: Set a random seed. Start from the month after first birth. Iterate from that month (call it month 1) to month 216 (216 months = 18 years). For each month, select the weeks within the month using NLSY crosswalk. Assign the modal status as the status in that month. If there is a tie, randomly choose one from the most popular statuses.

**Monthly Employment Sequences Construction for Pre-1978 Months for Main Sample - Python**

Code file: 2seq\_construction\_pre1978.ipynb

Output files:

* monthly\_df\_updated.csv

Content outline

* Goal: NLSY79 weekly statuses do not go back before 1978, but we have young mothers who gave births before 1978, so we will use work history raw variables to reconstruct the pre-1978 monthly statuses
* First, find mothers who have pre-1978 months with missing statuses – mothers who had 0 hour worked pre-1978 were assigned status 4 (nonemployed) for all pre-1978 months and they do not have missings
* Construct weekly statuses by finding out all the jobs that overlap with a given week (if overlap at least one day, consider R was working in the given week) and adding up all hours worked per week in all 5 jobs. Note: NLSY79 constructed total hours worked per week from 10 jobs, but only the first 5 jobs were released to the public and these are what we are using. If work hours >= 35, R worked full-time in the week. If work hours is between [0, 35), R worked part-time in the week. If hours is missing, R worked with missing hours. Then construct monthly statuses by finding the mode in the weekly statuses. If there is a tie, randomly pick one of the most frequent statuses – this is the same approach in the rest of the sequences
* Construct maternity leaves months. We don’t have within-job gap start/end dates. So our decision rule is: If R reported any nonemployed period prior to 1979 survey (29 cases) and R is working up to 1 month before birth or after 4 months after birth, then we assume R was on maternity leave in those months
* After the updating, we no longer have missing monthly statuses pre-1978, although we have 45 respondents with status 2 – working with missing hours

**Monthly Employment Sequences Construction for 14-year-seq sample - Python**

Code file: 2seq\_construction\_seq14.ipynb

Output files:

* monthly\_df\_seq14.csv: monthly statuses for 18 years after first childbirth

Content outline

* The main sample is part of the 14-year-seq sample. For the repeated respondents in the main sample, directly use their 18-year sequences and cut to 14-year. Repeat the same weekly status transformation and monthly status construction for the newly added respondents.

**Monthly Employment Sequences Construction Prebirth - Python**

Get additional one-year-long sequences prior to birth, for plotting purpose.

Code file: 2seq\_construction\_prebirth.ipynb

Output files:

* monthly\_df\_prebirth.csv: monthly statuses for one year prebirth

**Independent Variable Construction - Python**

Code file: 3IV\_construction.ipynb

Output file:

* iv\_df.csv
* missing\_marstat\_cases.csv
* marstat\_manual\_decisions.docx

|  |  |  |
| --- | --- | --- |
| Name | Label | Source |
| caseid\_1979 | Respondent ID |  |
| childbirth\_year | Birth year of first child | NLSY79 |
| Childbirth\_month | Birth month of first child |  |
| y\_dob | Respondent date of birth - year |  |
| m\_dob | Respondent date of birth - month |  |
| start\_y | The year in which the sequence starts | Assume birth happened in the last week of the month, and the sequence starts the month after |
| start\_m | The month in which the sequence starts |
| age | Age at first birth | Use both year and month to determine age at first birth |
| educ | Highest education achieved by one year prebirth | If first childbirth <= 1979, no yearly grade variable available:   * If age < 18 at one year prebirth, assume no high school * Otherwise, use grade reported in 1979   If first childbirth > 1979, use the highest grade reported by the time one year prebirth. Do not include grade 95 (i.e. ungraded study).  This is a categorical variable:   * 1-less than hs * 2-hs * 3-some col * 4-col * 5-advanced degrees |
| exp | Unadjusted work experience completed by one year prebirth | Add up all weekly hours completed by one year prior to birth (i.e. count up until the last week in the month (dob\_y-1, dob\_m)). If missing under 10%, expand to 100%. If missing more than 10%, assign missing and use multiple imputation later.  153 mothers had birth before 1978 and had worked before 1978. We are unable to determine whether work hours were logged before the prebirth time or after, so assign missing to them  Topcoded at 95th percentile. Bottomcoded at 5th percentile. Percentiles calculated using the first imputed dataset for only respondent who had worked at one year prior to birth |
| working | Whether R was currently working at one year prior to birth | 3 scenarios:  (1) Pre-birth work experience is zero: Set both dummies to zero    (2) One year pre-birth is before 1978 and we don't have job starting/ending date information prior to 1978 (119 cases):  Use the job dates reported in calendar Y/M/D. NLSY79 stores the original dates reported in calendar Y/M/D and also converts them into continuous week numbers. The latter is what we’ve been using when job dates do not exceed 1978.    (3) One year pre-birth is after 1978, use job dates to determine whether any job started before one year prebirth and any overlaped with one year prebirth |
| had\_job | Whether R had ever had any job one year prior to birth |
| marstat | Marital status one year prior to birth | Use marriage start/end dates to determine marital status. Assume marriage starts/ends in the first week of the month.  If undetermined after using marriage dates, use yearly marital status variables. Find out the month of interview in prebirth year. Then find out which are the two surrounding years. If the two surrounding years return the same martial status, use that as prebirth marital status.  If still undetermined, manually check cases.  20 missing remain.  Coding scheme:  0 – single  1 – married  2 – previously married  Problem: right now treat separated as status 2, but NLSY79 actually treat separated as married. Should we switch? |
| ma\_educ | Highest education completed by mother, reported in 1979 | 1 – never knew mom (-4) or less than hs  2 – hs  3 – some college  4 – college  5 – advanced  There are 216 missing cases. |
| ma\_ft | Whether mother worked full-time, reported in 1979 | 0 – mother did not work or worked part-time  1 – mother worked full-time  2 – no mother figure (101 cases)  missing (137 missings) |
| working35 | Whether R would like to work outside home at age 35 | Based on two questions:  exp\_1\_1979: WHAT WOULD R LIKE TO BE DOING AT AGE 35?   * 1 PRESENT JOB * 2 SOME OCCUPATION * 3 MARRIED, FAMILY * 4 OTHER (SPECIFY)   exp\_2\_1979: DOES R WANT TO WORK OUTSIDE HOME AT AGE 35?   * 1 YES * 0 NO   1yes – if R answered (1/2 in exp\_1\_1979), or (3 in exp\_1\_1979 and 1 in exp\_2\_1979)  0no – if R answered (4 in exp\_1\_1979), or (3 in exp\_1\_1979 and 0 in exp\_2\_1979)  Missing – 178 cases |
| womensroles | Family attitudes, reported in 1979. Higher value represents more traditional attitudes. | There are 8 gender roles questions. Picked 5 (using PCA):  1 - WOMAN'S PLACE IS IN THE HOME?  2 - WIFE WITH FAMILY HAS NO TIME FOR OTHER EMPLOYMENT?  4 - EMPLOYMENT OF WIVES LEADS TO JUVENILE DELINQUENCY?  6 - TRADITIONAL HUSBAND/WIFE ROLES BEST?  8 - WOMEN ARE HAPPIER IN TRADITIONAL ROLES?  Add up responses in the above question.  If only question has missing answer, expand the sum from the four questions to five. (101 cases)  If more than one missing, assign missing. (35 cases) |
| ideal | Ideal number of children, reported in 1979 | NLSY79 |
| race | 1 – Hispanic  2 – Black  3 – White | NLSY79 |
| numchild | Total fertility | Counted how many child dob years are valid |
| time\_btw\_birth | Number of months between first and second birth | None for Rs with only one child (579 cases) |
| yr\_cum\_exp | Yearly cumulative experience (in hours) within 18 years after first childbirth | Add up all weekly hours in 18 years.  If missing under <= 10%, expand multiple by 100. More than 10%, missing. (278 cases)  If missing more than 10%, assign missing (54 cases) |
| num\_ft\_months | Number of full-time month within 18 years after first childbirth | Missing months are not considered |
| num\_pt\_months | Number of part-time month within 18 years after first childbirth | Missing months are not considered |
| num\_nonemp\_months | Number of nonemployed month within 18 years after first childbirth | Missing months are not considered |
| num\_nonemp\_spells | Number of unemployment spells longer than 3 months | Missing months are not considered |

**Independent Variable Construction for 14-Year-Sequence Sample - Python**

Code file: 3IV\_construction\_seq14.ipynb

Output file:

* iv\_df\_seq14.csv

**Extra Independent Variable Construction for Post-1980-Birth Subsample - Python**

Code file: 3IV\_construction\_subsample.ipynb

Output file:

* iv\_df\_seq14.csv

Content outline:

* Keep only mothers with first birth after >= 1980. Sample size dropped from 343 to 2514.
* Construct IVs specific to this subsample

**Job traits: hourly rate and tenure**

### Scenario 1: no job is found (198 cases)

R had never worked before (198 cases). Hourly rate 0. Tenure 0. As expected, these respondents have 0 for the variable whether working at one year prebirth and 0 for the variable whehter ever held a job one year prebirth.

For all remaining respondents, drop repeated jobs. If a job lasted multiple years and were reported in multiple survey waves, track the same job and drop earlier references of the same job.

### Scenario 2: no overlap (459 cases)

Use the job that lasted the closest to the prebirth date. We previously chosen the job that started the latest by one year prebirth, but the job that lasted the longest is a better choice because this is the job that was active at the time closest to the point one year prebirth. For example, if a respondent had child in June 1989, her one year prebirth is June 1988. She had one job that started in January 1987 and ended in January 1988, and another job that started in February 1987 and ended in May 1988. We would want to pick the second job to retrieve hourlyr ate and tenure.

A small number of respondents have multiple jobs ending in the same week. Try the approaches in the following order to determine the job to retrieve job traits.

(1) Look for the job with the max weekly hours worked (0 case)

(2) Look for the job with the smallest job number (among jobs reported in the same year) (4 case)

### Scenario 3: one overlap (1690 cases)

If the overlaping job was reported in the same year as prebirth year, use it to retrieve job traits. Otherwise, try to trace back to the previous job reported <= prebirth year or closest to prebirth year, and use it to retrieve job traits.

### Scenario 4: multiple overlaps

* If there are two overlapping jobs,
  + If they were reported in different years, pick the one that was reported closer to prebirth year (9 cases)
  + otherwise, look for the job with the max weekly hours worked (14 case); or look for the job with the smallest job number (134 cases).
* If there are more than 2 overlapping jobs, print out their respondent id, prebirth year/month, and overlapping jobs information. First, look for the job with the max weekly hours worked (3 case). If still undecided, look for the job reported closer to prebirth year (7 cases).

Results: 204 cases missing hourly rates, 2 missing tenure

In order to know which year to apply the CPI inflation rate, I also reported three types of years – the year the hourly rate was reported, the year the job started, and the year the job ended. I think we should use the year the job ended as the year for CPI inflation rate. The job end year is always in the same year or before the year the hourly rate was reported. I think when respondents reported the hourly rate on a job, they are likely to recall the rate they earned at the end of the job, so the job end year is a better choice.

**Spouse wage**

If pre-birth marital status is single or previously married, spouse income is set to 0.

If pre-birth marital status is married, look for spouse wage reported in DOB year (because the spouse wage variable is defined to be TOTAL INCOME OF SPOUSE FROM WAGES AND SALARY IN PAST CALENDAR YEAR) and convert wage to 2016 dollar using DOB year -1 as the conversion year.

If spouse wage in DOB year is missing, use spouse wage reported the year before if it is valid. Convert wage to 2014 dollar using DOB year – 2 as the conversion year.

The CPI inflation calculator now has a month component. I chose the January of the various year and January of 2016.

|  |  |
| --- | --- |
| Spouse wage found in DOB survey year | 1096 |
| Spouse wage found in year before DOB survey year | 71 |
| Missing | 125 |

**Spouse hours worked per week**

Use same rule as above. Single and previously married respondents are set to have spouse hours of 0 and spouse hour group 0. However, 302 single respondents and 33 previously married respondents reported valid nonnegative spouse hours per week.

If found valid spouse hour in DOB year, use it. Otherwise, try spouse hours in the year before DOB year.

|  |  |
| --- | --- |
| Spouse hours found in DOB survey year | 1163 |
| Spouse hours found in year before DOB survey year | 0 |
| Missing | 129 |

**Multiple imputation and models**

Hourly rate, rate\_adj: bottomcoded at 5th percentile and topcoded at 95th percentile for the subsample who were currently working at one year prior to birth. Rate\_adj is set to missing for those who weren’t currently working.

Tenure: topcoded at 95th percentile for the subsample currently working at one year prior to birth. Tenure is set to missing for those who were not currently working at one year prior to birth.

Overall experience, exp: topcoded at 95th percentile for full sample and bottom code at 5th percentile. Using nonzero experience to find percentile cutpoints. Set exp to missing for those who had never had a job, run MI, and set their exp back to 0 afterwards. This way we will not get zero experience for those who had worked.

Spouse income, spwage\_adj: topcoded 95th percentile for the married subsample. Spouse income is set to missing for those single or previously married.

Spouse hours, sphour: set to missing for those single or previously married. Impute spouse hours and then categorize.