READ ME: Competition and Career Advancement

Julian Vedeler Johnsen, Hyejin Ku, Kjell Gunnar Salvanes

This document provides a description of the data, do-files, and variables used for the analyses in the article "Competition and Career Advancement" by Julian Vedeler Johnsen, Hyejin Ku, and Kjell G. Salvanes

Data Availability

The data used in the analyses are drawn from Norwegian administrative registers covering the full population of Norway. Due to confidentiality restrictions, we are unable to include our data in this replication package. With permission, the described data may be licensed from Statistics Norway. For detailed information (in English) on how to obtain this data, see https://www.ssb.no/en/data-til-forskning/utlan-av-data-til-forskere

All replication scripts are available at the following DOI: …

The data are drawn from the following registers:

- Birth register
- Social security registers
- Employee registers
- Tax registers
- Education registers

Do-files

Stata version 16 was used for the analyses. The do-files listed below produce the tables and figures in the paper. The individual tables and figures are found in the subfolders "Tables" and "Figures".

Item	Do-file	Dataset
Figure 1	fig1.do	Mainsample
Figure 2	N/A (figure is constructed without data)	
Figure 3	fig3.do	Mainsample
Figure 4	fig4.do	Mainsample
Figure 5	fig5.do	Mainsample
Figure 6	fig6.do	Mainsample
Figure A1	figA1.do	Mainsample
Figure A2	figA2.do	Mainsample
Figure A3	figA3.do	Mainsample
Table 1	tab1.do	Mainsample
Table 2	tab2.do	Mainsample
Table 3	tab3.do	Mainsample
Table 4	tab4.do	Mainsample
	tab4_row4.do	Donutsample
Table 5	tab5.do	Placebosample
Table 6	tab6.do	Mainsample
Table 7	tab7_panelA.do	Focalworkers_nonfathers
	tab7_panelBandC.do	Focalworkers_women
Table 8	tab8.do	Mainsample
Table A1	tabA1.do	Mainsample
Table A2	tabA2.do	Mainsample
Table A3	tabA3.do	Mainsample
Table A4	tabA4.do	Mainsample

Data sets and variables

Mainsample.dta

Population definition:

- All men in the Norwegian population registers who:
  - Fathered a first-born child in the window 01.07.1992 to 31.12.1993.
  - Was aged withing P10 to P90 of fathers of first-born children (age 25 to 36).
  - Worked in the same plant in the years -2, -1, and 0 relative to the child's birth year.
  - Had at least one male co-worker who had the same educational level, were within +/- 4 years of age, overlapped in the same plant in the years -2, -1, and 0 relative to the child's birth year, and who also fathered a child in the window 01.07.1992 to 31.12.1993.
  - Had earnings of at least 1G in each of the years -3 to 7 relative to the birth year of the child.

Panel structure:

- Annual panel data for years-since-birth -3 to 7 (calendar years 1989-2000)

Construction:

- To construct this dataset, we use the birth register to find all men who fathered a first-born child in the window 01.07.1992 to 31.12.1993. We exclude those who fathered multiple children (including twins) in the period. We further restrict the sample to those aged—at child's birth—within the 10th and the 90th percentiles (25 and 36 years) of the age distribution. We define these as focal fathers.
- We then merge these data to the employee registers and further restrict the sample to those who had the same plant identifier in years -2, -1, and 0 relative to the birth year of the child.
- Within the employee registers we find and define competitors as all male co-workers who had the same educational level as the focal fathers, were within +/- 4 years of age of the focal father, overlapped with the focal father in the same plant in the years -2, -1, and 0 relative to the birth year of the child of the focal father, and who also fathered a child in the window 01.07.1992 to 31.12.1993. To enable our analysis, we restrict the sample to focal fathers who had at least one such competitor.
- We then link this dataset of focal fathers and competitors to social security registers containing data on parental leave take-up.
- Finally, we use this population to construct panel data and merge in outcomes from the tax registers. To maintain a balanced sample, we restrict the sample to focal fathers with earnings of at least 1G in each of the panel data periods.

Donutsample.dta

- This sample is identical to Mainsample.dta, except that we exclude all contests that include any births (to focal father or competitor) occurring in March of April 1993.

Time-invariant variables in Mainsample.dta

Variable name	Description
npid	Personal identifier
t	Years since calendar year of child's birth
year	Calendar year
Z	Indicator for birth occurring after April 1st, 1993
D	Indicator for taking up parental leave (=1 if days of leave>0)
S	Share of competitors eligible for paid paternity leave (child born after April 1st, 1993)
C	Share of competitors taking paid paternity leave
leave_duration	Days of parental leave
nmb_competitors	# of competitors
contest_size	Size of contest (# of competitors + 1)
plant_0	Identifier of plant worked in at calendar year of child's birth
employees	Firm size of plant_0 (# of registered employees)
micro	Indicator for firm size <10
small	Indicator for firm size>=10 & <50
medium	Indicator for firm size>=50 & <250
public	=1 if plant is in public sector (based on industry codes)
mean_income	average earnings in the plant
mean_age	average age of employees in the plant
sharemale	Share of male employees in the plant
shareHS	Share of employees in the plant with >= a high school degree
mean_eduy	Average years of education of employees in the plant
sd_earnings_growth_r	The standard deviation of earnings growth from (focal father's) YSB -2 to 0 of male workers aged 25 to 36 who are employed in the same plant as the focal father
growth	Percentage growth in firm size of focal father's plant from (focal father's) YSB-2 to 0.
share_above_p50	Share of top earners within the focal father's plant, with top earners defined as workers with an above-median wage, relative to a wage distribution calculated within a county by two-digit industry sector by firm-size category cell
byr_child	Calendar year of child's birth
bmonth_child	Calendar month of child's birth
male_child	=1 if sex of child is male
age_0	Age per 31.12 of calendar year of child's birth
edulevel_0	Education level attained per 31.12 of calendar year of child's birth (1: <= 2 years of high-school, 2: high-school diploma, 3: <= 4 years of college, 4: >4 years of college)

Time-variant variables in Mainsample.dta

Variable name	Description
income	earnings (inflation-adjusted to 1993 NOK)
ln_income	log earnings (inflation-adjusted to 1993 NOK)
FD_ln_income	log earnings relative to log earnings in years-since-birth = -1
firm_change	Indicator variable equal to 1 if the plant identifier in a given year is not equal to plant_0 (0 otherwise)
age	age per 31.12 of calendar year

Placebosample.dta

Population definition:

- All men in the Norwegian population registers who:
    o Fathered a first-born child in the window 01.07.1991 to 31.12.1992.
    o Was aged withing P10 to P90 of fathers of first-born children (age 25 to 36).
    o Worked in the same plant in the years -2, -1, and 0 relative to the child's birth year.
    o Had at least one male co-worker who had the same educational level, were within +/- 4 years of age, overlapped in the same plant in the years -2, -1, and 0 relative to the child's birth year, and who also fathered a child in the window 01.07.1991 to 31.12.1992.
    o Had earnings of at least 1G in each of the years -3 to 7 relative to the birth year of the child.

Panel structure:

- Annual panel data for years-since-birth -3 to 7 (calendar years 1988-1999)

Construction:

- To construct this dataset, we use the birth register to find all men who fathered a first-born child in the window 01.07.1991 to 31.12.1992. We exclude those who fathered multiple children (including twins) in the period. We further restrict the sample to those aged—at child's birth—within the 10th and the 90th percentiles (25 and 36 years) of the age distribution. We define these as focal fathers.
- We then merge these data to the employee registers and further restrict the sample to those who had the same plant identifier in years -2, -1, and 0 relative to the birth year of the child.
- Within the employee registers we find and define competitors as all male co-workers who had the same educational level as the focal fathers, were within +/- 4 years of age of the focal father, overlapped with the focal father in the same plant in the years -2, -1, and 0 relative to the birth year of the child of the focal father, and who also fathered a child in the window 01.07.1991 to 31.12.1992. To enable our analysis, we restrict the sample to focal fathers who had at least one such competitor.
- Finally, we use this population to construct panel data and merge in outcomes from the tax registers. To maintain a balanced sample, we restrict the sample to focal fathers with earnings of at least 1G in each of the panel data periods.

Time-invariant variables in Placebosample.dta

Variable name	Description
npid	Personal identifier
t	Years since calendar year of child's birth
year	Calendar year
Z	Indicator for birth occurring after April 1st, 1992
S	Share of competitors eligible for placebo reform (child born after April 1st, 1992)
plant_0	Identifier of plant worked in at calendar year of child's birth
byr_child	Calendar year of child's birth
bmonth_child	Calendar month of child's birth
age_0	Age per 31.12 of calendar year of child's birth
edulevel_0	Education level attained per 31.12 of calendar year of child's birth

Time-variant variables in Placebosample.dta

Variable name	Description
ln_income	log earnings (inflation-adjusted to 1993 NOK)
age	age per 31.12 of calendar year

Focalworkers_nonfathers.dta

Population definition:

- All men in the Norwegian population registers who:
    o Did not father a child in the window 01.07.1992 to 31.12.1993.
    o Were aged 25 to 36 in 1992.
    o Worked in the same plant in the calendar years 1990 to 1993.
    o Had at least two male co-workers in his plant who had the same educational level, were within +/- 4 years of age, overlapped in the same plant in the years 1990 to 1993, and who fathered a child in the window 01.07.1992 to 31.12.1993.
    o Had earnings of at least 1G in each of the years 1990 to 1999.

Panel structure:

- Annual panel data for years 1990 to 1999.

Construction:

- To construct this dataset, we start by selecting all plants who, among the employees listed for 1992, had at least two men who fathered a child in the window 01.07.1992 to 31.12.1993. We exclude plants in which at least one male employee fathered multiple children (including twins) in the period.
- Within these plants, we select male employees with stable employment in the plant in the period 1990 to 1993 and who were aged 25 to 36 in 1992. We exclude men who fathered a child in the window 01.07.1992 to 31.12.1993.
- For these focal non-fathers, we find and define competitors as all male co-workers who had the same educational level as the focal non-father, were within +/- 4 years of age of the focal non-father, overlapped with the focal non-father in the same plant in the years 1990 to 1993, and who fathered a child in the window 01.07.1992 to 31.12.1993. To enable our analysis, we restrict the sample to focal non-fathers who had at least one such competitor.
- We use this population to construct panel data and merge in outcomes from the tax registers. To maintain a balanced sample, we restrict the sample to focal non-fathers with earnings of at least 1G in each of the panel data periods.

Time-invariant variables in Focalworkers_nonfathers.dta

Variable name	Description
npid	Personal identifier
year	Calendar year
S	Share of competitors eligible for paid paternity leave (child born after April 1st, 1993)
plant_1992	Identifier of plant worked in 1992
age92	Age per 31.12 of 1992
edulevel_0	Education level attained per 31.12 of 1992 (1: <= 2 years of high-school, 2: high-school diploma, 3: <= 4 years of college, 4: >4 years of college)

Time-variant variables in Focalworkers_nonfathers.dta

Variable name	Description
income	earnings (inflation-adjusted to 1993 NOK)
ln_income	log earnings (inflation-adjusted to 1993 NOK)
age	age per 31.12 of calendar year

Focalworkers_women.dta

Population definition:

- All women in the Norwegian population registers who:
  - Were aged 25 to 36 in 1992.
  - Worked in the same plant in the calendar years 1990 to 1993.
  - Had at least two male co-workers in her plant who had the same educational level, were within +/- 4 years of age, overlapped in the same plant in the years 1990 to 1993, and who fathered a child in the window 01.07.1992 to 31.12.1993.
  - Had earnings of at least 1G in each of the years 1990 to 1999.

Panel structure:

- Annual panel data for years 1990 to 1999.

Construction:

- To construct this dataset, we start by selecting all plants who, among the employees listed for 1992, had at least two men who fathered a child in the window 01.07.1992 to 31.12.1993. We exclude plants in which at least one male employee fathered multiple children (including twins) in the period.
- Within these plants, we select female employees with stable employment in the plant in the period 1990 to 1993 and who were aged 25 to 36 in 1992.
- For these focal women, we find and define competitors as all male co-workers who had the same educational level as the focal woman, were within +/- 4 years of age of the focal woman, overlapped with the focal woman in the same plant in the years 1990 to 1993, and who fathered a child in the window 01.07.1992 to 31.12.1993. To enable our analysis, we restrict the sample to focal women who had at least one such competitor.
- We use this population to construct panel data and merge in outcomes from the tax registers. To maintain a balanced sample, we restrict the sample to focal women with earnings of at least 1G in each of the panel data periods.

Time-invariant variables in Focalworkers_women.dta

Variable name	Description
npid	Personal identifier
year	Calendar year
S	Share of competitors eligible for paid paternity leave (child born after April 1st, 1993)
plant_1992	Identifier of plant worked in 1992
age92	Age per 31.12 of 1992
edulevel_0	Education level attained per 31.12 of 1992 (1: <= 2 years of high-school, 2: high-school diploma, 3: <= 4 years of college, 4: >4 years of college)
birth	= 1 if the woman had a birth in the window 01.07.1992 to 31.12.1993

Time-variant variables in Focalworkers_women.dta

Variable name	Description
income	earnings (inflation-adjusted to 1993 NOK)
ln_income	log earnings (inflation-adjusted to 1993 NOK)
age	age per 31.12 of calendar year

Stata ado-files used:

- reghdfe is used in tab8.do and tabA4.do.
  To install, use the following command: ssc install reghdfe