# Replication Package

*"Childbirth and Firm Performance: Evidence from Norwegian Entrepreneurs"*

John Bonney, Luigi Pistaferri, and Alessandra Voena

---

## Overview

This replication package contains all code needed to reproduce the tables and figures in the paper, starting from the raw data described below. The analysis uses Norwegian administrative register data (1995–2019), European Union Statistics on Income and Living Conditions (EU-SILC) microdata (2004–2023), and O*NET work activity data. The main analysis is conducted in Stata. Tables and figures are assembled in R, and the O*NET data are linked to Norwegian industries in Python.

## Data Availability

### 1. Norwegian Administrative Register Data (Restricted Access)

The primary data employed in the analysis are drawn from Norwegian administrative registers maintained by Statistics Norway. Statistics Norway provides micro data for research projects, either by granting access after application and approvals to anonymized data, or as part of a Norwegian research project. For details, see https://www.ssb.no/en/data-til-forskning.

Inquiries about access to data from Statistics Norway should be addressed to: **mikrodata@ssb.no**.

All datasets are generated from the raw files of the Norwegian administrative registers. The scripts generating these datasets can be provided upon request. The scripts are written as if this data were to be placed in `data/admin/raw` (in practice, they are analyzed on a secure server maintained by Statistics Norway).

| Dataset | Description | Years |
|---|---|---|
| bs9519.dta | Firm balance sheet data (revenues, costs, profits, assets, employment, NACE codes, founding dates) | 1995–2019 |
| ee_extended_9518.dta | Employer-employee register (wages, hours, occupation, industry, municipality), person level | 1995–2018 |
| founding_dates.dta | Firm founding dates | –2019 |
| firm_ownership_01_19.dta | Individual ownership shares in private non-financial LLCs | 2001–2019 |
| ancestry_as_of_2021.dta | Birth register linking children to parents | –2021 |
| living_64_20.dta | Municipality of residence | –2020 |
| lnr_mlnr_flnr_updated2013.dta | Parent-child linkage | –2013 |

| | | |
|---|---|---|
| constant_traits.dta | Time-invariant individual characteristics (gender, birth year) | –2013 |
| education_nus2000.dta | Education register (NUS2000 classification) | — |
| marital_cohabit_91_20.dta | Marital status | 1991–2020 |
| increg_93_19.dta | Income register | 1993–2019 |

#### 2. EU-SILC Microdata (Restricted Access)

Cross-sectional EU-SILC microdata for Norway (2004–2023) are obtained from Eurostat. Access requires a research contract with Eurostat; see https://ec.europa.eu/eurostat/web/microdata/european-union-statistics-on-income-and-living-conditions.

The following files are used (one per year, for years 2004–2023):

- UDB_cNO[YY]P.csv — Personal data (P-file): labor status, hours worked, self-reported health, income, education, NACE codes
- UDB_cNO[YY]R.csv — Personal register (R-file): demographics, household composition, childcare
- UDB_cNO[YY]H.csv — Household data (H-file): child/family allowances

Raw CSV files should be placed in `data/silc/Raw_Data/NO_CROSS/[YYYY]/`.

#### 3. O*NET Data (Public)

O*NET work activity and occupation data are publicly available from https://www.onetcenter.org/database.html. The version used in this project is 28.0. The files used are:

- Work Context.txt
- Occupation Data.txt

These files are included in `data/onet/`.

#### 4. Industry Crosswalk (Constructed)

The crosswalk between Norwegian SIC industry codes and O*NET occupations is constructed by the scripts `match_industry_to_occ.py` and `occ_crosswalk.R`. Supporting crosswalk files are in `data/crosswalk/`.

## Instructions for Replicators

First, update the project root in `code/master.do`. Then, take the following steps:

1. Build the crosswalk from O*NET to Norwegian industries, running:
    a. `code/admin/data_construction/match_industry_to_occ.py`
    b. `code/admin/data_construction/occ_crosswalk.R`

2. Run the master file: `code/master.do` (see below for details)
3. Produce the exhibits: `code/exhibits.R`

## Program Descriptions

Programs are listed in execution order as called by `master.do`.

**Part 1: Data Construction**

| # | Program | Description |
|---|---|---|
| 1 | identify_births.do | Links children in the birth register to their parents; identifies the timing of first births for the sample period. |
| 2 | grandparent_location.do | Determines whether grandparents live in the same municipality as the entrepreneur, using the parent-child linkage and municipality of residence files. |

**Part 2: Panel Construction**

| # | Program | Description |
|---|---|---|
| 3 | setup_panel.do (arg: 33) | Constructs the main analysis panel using a one-third ownership threshold. Merges firm, individual, and family characteristics; constructs outcome variables; creates stacked event study cohorts around first birth. |
| 4 | setup_panel.do (arg: 25) | Constructs the robustness panel using a one-quarter ownership threshold. |

**Part 3: Analysis**

| # | Program | Description | Exhibits |
|---|---|---|---|
| 5 | summary_stats.do | Computes summary statistics for the estimation sample. | Table 1 |
| 6 | estimation_main.do | Estimates stacked difference-in-differences event studies for all main outcomes (profits, revenues, costs, employment, ownership, wages). | Figures 1, A1; Tables 2, A4 |
| 7 | event_study_ability.do | Estimates profit event studies separately by quartile of pre-birth firm | Figures 2, A2 |

| | | | |
|---|---|---|---|
| | | fixed effect, using both 2-period and 3-period pre-birth windows. | |
| 8 | estimation_heterogeneity.do | Estimates profit event studies split by five dimensions: any employees, daily face-to-face contact (O*NET), proximity to grandparents, majority ownership, and partner employment. | Tables 3, A5 |

**Part 4: Robustness**

| # | Program | Description | Exhibits |
|---|---|---|---|
| 9 | robustness.do | Re-estimates main specifications under alternative sample criteria (firm age, ownership thresholds), control specifications (2-digit industry, education, parent controls), and trimming thresholds (5%, 25%). | Tables A1, A2, A3 |
| 10 | wage_penalty.do | Estimates child penalty in wages for non-entrepreneurial wage earners as a comparison. | Figure A3 |

**Part 5: EU-SILC Analysis**

| # | Program | Description | Exhibits |
|---|---|---|---|
| 11 | cleaning_P_vars.do | Cleans personal interview data (labor status, hours, health, income, education) for all EU-SILC years. | — |
| 12 | cleaning_R_vars.do | Cleans personal register data (demographics, household composition, childcare) and constructs parent-child linkages within households. | — |
| 13 | cleaning_h_vars.do | Cleans household-level data (child/family allowances). | — |
| 14 | analysis.do | Merges cleaned EU-SILC files; estimates | Figure 3, Figure A4, Table A6 |

| | | | |
|---|---|---|---|
| | | regressions of hours worked and self-reported health on child age indicators for self-employed parents. | |

**Exhibits**

| # | Program | Description |
|---|---|---|
| 15 | exhibits.R | Reads all estimation output from estimates/; produces final tables (.tex) and figures (.pdf) in exhibits/. |