Job Transitions and Employee Earnings After Acquisitions: Linking Corporate and Worker Outcomes

David Arnold (daarnold@uscd.edu)[1]
Kevin Milligan (kevin.milligan@ubc.ca)[2]
Terry (Seok Min) Moon (tsmoon@mail.ubc.ca)[3]
Amirhossein Tavakoli (ah.tavakoli@ubc.ca)[4]

RESTAT - 31581: README file for Code and Data Availability

1 Overview

This ReadMe file describes how to replicate the analysis and exhibits in "Job Transitions and Employee Earnings After Acquisitions: Linking Corporate and Worker Outcomes." The analysis uses administrative data from the Canadian Employer–Employee Dynamics Database (CEEDD) at the Canadian Centre for Data Development and Economic Research (CDER), Statistics Canada. Because the underlying data are confidential and subject to the *Statistics Act*, they cannot be shared publicly and the paper has been exempted from the data requirement. Researchers wishing to replicate or extend our results must obtain independent access to the same data sources through Statistics Canada. Detailed guidance on accessing the data is provided below.

2 Obtaining the administrative data

All empirical analyses in the paper must be performed at a Research Data Centre (RDC) in Canada, which is part of the Canadian Research Data Centre Network (CRDCN) providing restricted access to confidential micro-level data. The data used in the analyses will be preserved at the RDC for the next five years. To access the data, researchers must submit a research proposal for review and approval by Statistics Canada. More details on how to submit a research proposal can be found at:

https://www.statcan.gc.ca/en/microdata/data-centres/access

Alternatively, researchers may request permission from Statistics Canada by emailing the data access division (statcan.dad-apu-dad-uta.statcan@statcan.gc.ca) with the project number (6730); upon approval, they can access the raw data files at the RDC for the purpose of replication only.

The RDC is available at the headquarters of Statistics Canada in Ottawa, ON, and at multiple universities across Canada.

[1] Department of Economics, University of California San Diego
[2] Vancouver School of Economics, University of British Columbia
[3] Vancouver School of Economics, University of British Columbia
[4] Vancouver School of Economics, University of British Columbia

The data files used in this project are:

Dataset	Description	Key variables (some examples)
T4ROE	T4 slips and Records of Employment	Worker earnings, employer ID
T1	Individual income tax returns	Worker age, sex, UI earnings
NALMF	National Accounts Longitudinal Microdata File	Firm revenue, employment, payroll, assets
T2 (Schedule 50)	Corporate shareholder records	Owner identity, realized capital gains
SDC Platinum (linked to NALMF)	M&A deal database	Deal year, form (merger/acquisition), acquirer/target flag, industry
canada_cpi	National Consumer Price Index (CPI)	CPI from 2001 to 2017

SDC Platinum has near-universal coverage of M&A transactions for publicly traded and private companies in Canada. The M&A records from this source were linked to NALMF firm identifiers using all available identifying variables (firm name, NAICS code, address). The match rate is approximately 75 percent on average from 2001 to 2017. The final data are separated into two parts:

1. mna_linked_to_nalmf.dta: firm-deal level records of mergers and acquisitions, including deal form (acquisition, merger), deal year, and identifiers for target and acquiring firms.
2. mna_linked_to_nalmf_with_naics.dta: firm-deal level records including NAICS industry codes of the target and acquiring firms.

The following information from external sources is used to deflate price-related variables, National Consumer Price Index (CPI) from 2001 to 2017:

https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1810000501

The annual CPI series is saved as canada_cpi.dta in the project folder.

All firm and worker identifiers in the confidential microdata are synthetic identifiers assigned by Statistics Canada and do not correspond to real entity identifiers.

3 Software Requirements

Analysis was performed in **Stata 17** running on [Windows 11 Pro 25H2].

Hardware used for the last replication:

- Processor: AMD EPYC 7742 64-Core
- Memory: 128 GB RAM
- Computing environment: single local machine (no cluster or distributed computing)
- Storage requirement: 200 GB

The following user-written Stata packages are required. Within the RDC/CDER environment, these are stored in a local `ado` directory (path set via `sysdir set PLUS` in `master.do`) and do not require internet access. To install them in a connected environment:

```
ssc install egenmore       // extended egen functions
ssc install fillin         // fill in panel observations
ssc install missings       // utilities for missing values
ssc install ftools         // fast Stata commands
ssc install gtools         // fast alternatives to collapse, egen, etc.
gtools, upgrade
ssc install coefplot       // coefficient and event-study plots
ssc install reghdfe        // two-way fixed-effect estimation
ssc install group2hdfe     // AKM group identifier (used with reghdfe)
```

4 Folder Directory

The replication package consists of the do-files listed in Section 5. Because the raw data and processed datasets reside within the secure RDC/CDER environment, the package does not include a data folder. All output (tables and figures) is written to the directory specified by the global `$out` in `master.do`.

The recommended directory structure within the RDC project is:

```
replication_package/           <- $code (set in master.do)
├── README.md
├── README.pdf
├── master.do
├── functions.do
└── [all other do-files]

[elsewhere on RDC server]
├── data/                      <- $data  (intermediate datasets)
├── output/                    <- $out   (tables and figures)
└── [project folder]/
    ├── PROJECT_6730/           <- $data0
    └── PROJECT_6730_10138/     <- $data1, $data2, $data3
```

Before running, update the six global directory paths at the top of `master.do`:

Global	Contents
`$code`	Location of the replication package (do-files)
`$out`	Destination for output tables and figures
`$data`	Processed and intermediate datasets
`$data0`	Raw M&A–NALMF linked file
`$data1`	Raw T4ROE, NALMF, and T1 files
`$data2`	Supplementary firm and worker variables
`$data3`	Geographic identifiers (commuting zones)

5 Do-Files Directory

Table 1 describes each Stata do-file, its purpose, and the exhibits (tables and figures) it generates. Run `master.do` from top to bottom to reproduce all results. The do-files must be executed in the order listed; each file depends on datasets produced by earlier files.

Table 1. Description of Replication Scripts and Generated Datasets/Exhibits

Script	Description	Datasets/Exhibits generated
`master.do`	Master do-file. Sets global directory paths, Stata settings, and calls all other do-files in sequence. Update the six path globals before running.	—
`functions.do`	Defines the `make_eventstudy` program, a `coefplot` wrapper used to produce event-study plots in all analysis do-files. Defines the `make_eventstudy_decomp` program used to produce decomposition plot.	—
Data Cleaning		
`clean_mna.do`	Builds the deal-level M&A dataset from SDC Platinum linked to NALMF. Retains completed deals; drops share buybacks and self-deals.	`all_mna.dta`, `first_mna.dta`, `repeat_acquirer.dta`
`worker_firm_panel.do`	Builds the worker–firm panel, moonlighting indicator, and firm employment panel from T4ROE records (2001–2017).	`worker_firm_panel.dta`, `moonlighter.dta`, `emp_panel.dta`
`clean_firm.do`	Creates the annual firm-level dataset for propensity-score matching from NALMF, merged with M&A status, AKM employer fixed effects, and match effects.	`firm_YYYY.dta` (one file per year, 2001–2017), `firm_level_emp.dta`
`clean_mna_firm.do`	Creates a panel of all M&A firms (acquirers and targets) over time. Used in constructing Figure 1.	`NALMF_mna.dta`
`akm_estimation.do`	Estimates employer fixed effects following Abowd, Kramarz, and Margolis (1999). Restricted to movers in the largest connected set; firms with ≥ 5 employees; real earnings ≥ CAD 3,900.	`akm.dta`
`match_effect_estimation.do`	Estimates worker–firm match effects following Woodcock	`match_effect.dta`

	(2008), controlling for tenure within each worker–firm match.	
clean_worker.do	Creates the annual worker-level dataset for matching. Merges T4ROE worker records with firm characteristics, commuting zone identifiers, AKM effects, and match effects. Also constructs auxiliary datasets for market-overlap and concentration analysis.	worker_YYYY.dta, firm_year_cz.dta, same_market.dta, same_market_firm.dta, treated_market.dta
concentration.do	Computes labor market concentration measures at the commuting zone × 4-digit NAICS level: standard HHI (employment-share based) and generalized flow-adjusted HHI (GHHI) following Arnold (2025).	concentration.dta, ghhi.dta
clean_cap_gain.do	Builds the owner capital gains dataset from T2 Schedule 50 shareholder records. Merges with the matched and eligible firm lists. **Note:** must be run after firm-level matching (matching_firm.do and matching_firm_eligible.do).	firm_owner.dta, eligible_firm_owner.dta
Matching		
matching_firm.do	Baseline firm propensity-score matching: 1:1 nearest-neighbor without replacement within a caliper, exact cell matching on 2-digit NAICS, province, and binned covariates (revenue, payrolls, age). Constructs the event-study panel of matched firms.	firm_matched_list.dta, firm_matched.dta
matching_worker.do	Baseline worker propensity-score matching within cells defined by sector × province × sex × 5-year age bin. Constructs the event-study panel of matched workers.	worker_matched_list.dta, worker_matched.dta
matching_worker_w_tenure.do	Robustness: adds worker tenure as an additional matching covariate.	worker_matched_w_tenure.dta
matching_worker_w_wage.do	Robustness: adds the worker's within-firm earnings distribution percentile as a	worker_matched_w_wage.dta

	matching variable.	
matching_firm_diff_market.do	Robustness: requires matched firm pairs to operate in different geographic markets (different CZ × industry cells).	firm_matched_diff_market.dta
matching_worker_diff_market.do	Robustness: matched worker pairs drawn from firms in different geographic markets.	worker_matched_diff_market.dta
matching_firm_w_roa.do	Robustness: adds return on assets (ROA) as an additional firm-level matching variable.	firm_matched_w_roa.dta
matching_worker_w_roa.do	Robustness: constructs the worker event-study panel within ROA-matched firm pairs.	worker_matched_w_roa.dta
matching_firm_eligible.do	Constructs the unmatched eligible firm sample: firms that passed the selection criteria but were not matched. Used in balance-table comparisons.	firm_eligible_list.dta
matching_worker_eligible.do	Constructs the unmatched eligible worker sample.	worker_eligible_list.dta
Analysis		
descriptive_stats.do	Produces the M&A deal count figure, matched-sample summary statistics, characteristics of unmatched eligible firms and workers, AKM exogenous mobility test, exit and unemployment transition probabilities, and propensity score overlap figures.	Figure 1, Table 1, Table A9, Figure A9, Figure A11 (panel A), Figure A12
analyze_main.do	Estimates firm-level difference-in-differences event-study results for main outcomes (employment, average payrolls, profit margins, ROA) and extended firm outcomes (revenue, markups, NIBTEI per worker, capital gains).	Table 2, Figure 2, Table B6, Figure B6
analyze_worker.do	Estimates worker-level difference-in-differences event-study results for earnings, job transitions, changes in average firm characteristics of job movers, and heterogeneity by tenure and earnings quintile.	Table 3–6, Figure 3-6

analyze_robust.do	Produces all Appendix A robustness checks: alternative standard error clustering, alternative matching specifications, outcomes in levels, by one-time vs. repeat acquirers, private-firms-only sample, ROA matching, tenure matching, unmatched eligible firms and workers inclusion, and AKM wage decomposition into employer fixed effect, match effect, and residual channels. Also produces Figure A11 (panel B).	Table A1–A8, Figure A1–A8, Figure A10, Figure A11 (panel B)
analyze_hetero.do	Produces all Appendix B heterogeneity analyses: by local vs. national M&As, acquisition vs. merger type, market overlap, initial HHI level, initial GHHI level, tradable vs. non-tradable sectors, between- vs. within-industry job movers, voluntary vs. involuntary separation, medium-tenure group, and age heterogeneity.	Table B1–B5, B7–B12, Figure B1–B5, B7–B12

6 Variable Definitions

The tables below describe the key variables from each source dataset and any derived variables constructed from them. Note that other auxiliary variables are defined within their corresponding code. All firm and worker identifiers are synthetic and assigned by Statistics Canada.

Table 2. NALMF (firm-level, one file per year: `firm_YYYY.dta`)

Original Variable	Description	Derived Variable	Description
entid_syn	Synthetic firm identifier	—	—
year	Fiscal year	—	—
naics	4-digit NAICS industry code	naics2	2-digit NAICS (grouped sector)
OPAddressProvince	Province of firm's postal address	—	Used as matching cell variable
sac_syn	Statistical Area Classification code	—	Commuting zone (CZ) proxy
BirthDate	Firm registration date	age	Firm age (years)
total_revenue	Total revenue	—	Matching variable and outcome

total_expenses	Total expenses	netprof	(Revenue − Expenses) / Revenue
total_assets	Total assets	roa	(Revenue − Expenses) / Assets
PD7_AvgEmp_NonZero	Average number of employees	avg_wage	Main firm-level outcome
T4_Payroll	Total Payrolls	avg_wage, avg_payrolls	Main firm-level outcome
net_income_befor_taxextraitems	Net income before taxes and extraordinary items	ebit_lbr	Profitability per worker (Appendix B)

Table 3. T4ROE (worker–employer records, one file per year: worker_YYYY.dta)

Original Variable	Description	Derived Variable	Description
casenum2019	Synthetic worker identifier	—	—
entid_syn	Synthetic employer identifier	—	Links to NALMF
year	Tax year	—	—
t4earn	Pre-tax earnings on T4 slip	log(Earnings)	Main worker-level outcome
moved	Transition Indicator	Moved	Main worker-level outcome
moved_ind	=1 if a worker transition into a different industry	—	—

Table 4. T1 (individual income tax returns, merged into worker file)

Original Variable	Description	Derived Variable	Description
casenum2019	Synthetic individual identifier	—	Links to T4ROE
t1_sex_recorded	Sex of worker	male	1 = male; used in matching cell
t1_age_recorded	Age of worker	age_bin	5-year age bin; used in matching cell
cap_tot_net_calc	Total Realized Capital	g	Total realized capital gains for all owners of the firm

Table 5. T2 Schedule 50 (shareholder records, used in clean_cap_gain.do)

Original Variable	Description	Derived Variable	Description
casenum2019	Synthetic owner	—	—

	identifier		
entid_syn	Synthetic identifier of owned firm	—	Links to NALMF
year	Year	—	—

Table 6. SDC Platinum (deal-level, linked to NALMF in clean_mna.do)

Original Variable	Description	Derived Variable	Description
DEAL	SDC Platinum deal identifier	—	—
Acquirer	1 = acquirer firm; 0 = target firm	—	—
Status	'C' = Completed	—	—
Form	1 = Asset acquisition (Acq. of Assets / Acq. of Certified Assets) 2 = Interest acquisition (Majority, Partial, or Remaining interest) 3 = Merger	matched_merger	—
treated	1 = M&A firm; 0 = control	—	Treatment indicator
within_sector	= 1 if acquirer and target share the same 2-digit NAICS	—	—
within_all	= 1 if acquirer and target share the same 4-digit NAICS	—	—
num_deal	Number of Deal by Firm ID	—	—

Table 7. Constructed variables (AKM estimation and match effects)

Variable	Description
fe	AKM employer fixed effect: log wage premium for firm j, estimated from akm_estimation.do
logearnings_met	Worker–firm match effect estimated following Woodcock (2008), from match_effect_estimation.do
hhi	Standard Herfindahl–Hirschman Index based on employment shares within CZ × NAICS markets
ghhi	Generalized (flow-adjusted) HHI following Arnold (2025)
year_prior	Year before the M&A event
t	Event time (years relative to deal)
ds_{k}	Event-time indicator

pairid	Matched pair identifier
bin_tenure	2-year tenure bins (for matching and heterogeneity analysis)
bin_t4earn	Within-firm wage quintile (for matching and heterogeneity analysis)
move_type	=0 if worker never moves, =1 if the worker moves within the target firm, =2 if worker moves to the acquiring firm, =3 if worker moves to other firms

7 List of Outputs (Exhibits)

All output files are saved to the directory specified by `$out` in `master.do`. Tables are saved as `.csv` files and figures as `.pdf` files.

Table 8. Main Text Exhibits

Exhibit	Script	Output file(s)
Figure 1	descriptive_stats.do	Figure1.pdf
Table 1	descriptive_stats.do	table1.csv
Table 2	analyze_main.do	table2.csv
Figure 2 A–H	analyze_main.do	Figure2_A.pdf, ..., Figure2_H.pdf
Table 3	analyze_worker.do	table3.csv
Figure 3 A–C	analyze_worker.do	Figure3_A.pdf, Figure3_B.pdf, Figure3_C.pdf
Table 4	analyze_worker.do	table4.csv
Figure 4 A–D	analyze_worker.do	Figure4_A.pdf, ..., Figure4_D.pdf
Table 5	analyze_worker.do	table5.csv
Figure 5 A–D	analyze_worker.do	Figure5_A.pdf, ..., Figure5_D.pdf
Table 6	analyze_worker.do	table6.csv
Figure 6 A–D	analyze_worker.do	Figure6_A.pdf, ..., Figure6_D.pdf

Table 9. Appendix A Exhibits

Exhibit	Script	Output file(s)
Table A1	analyze_robust.do	tableA1.csv
Figure A1 A-C	analyze_robust.do	FigureA1_A.pdf, FigureA1_B.pdf, FigureA1_C.pdf
Table A2	analyze_robust.do	tableA2.csv
Figure A2	analyze_robust.do	FigureA2_A.pdf, FigureA2_B.pdf,

A–C		FigureA2_C.pdf
Table A3	analyze_robust.do	tableA3.csv
Figure A3 A–C	analyze_robust.do	FigureA3_A.pdf, FigureA3_B.pdf, FigureA3_C.pdf
Table A4	analyze_robust.do	tableA4.csv
Figure A4 A–B	analyze_robust.do	FigureA4_A.pdf, FigureA4_B.pdf
Table A5	analyze_robust.do	tableA5.csv
Figure A5 A–C	analyze_robust.do	FigureA5_A.pdf, FigureA5_B.pdf, FigureA5_C.pdf
Table A6	analyze_robust.do	tableA6.csv
Figure A6 A–C	analyze_robust.do	FigureA6_A.pdf, FigureA6_B.pdf, FigureA6_C.pdf
Table A7	analyze_robust.do	tableA7.csv
Figure A7 A–C	analyze_robust.do	FigureA7_A.pdf, FigureA7_B.pdf, FigureA7_C.pdf
Table A8	analyze_robust.do	tableA8.csv
Figure A8 A–C	analyze_robust.do	FigureA8_A.pdf, FigureA8_B.pdf, FigureA8_C.pdf
Figure A9	descriptive_stats.do	FigureA9.pdf
Table A9	descriptive_stats.do	tableA9.csv
Figure A10	analyze_robust.do	FigureA10.pdf
Figure A11 (panel A)	descriptive_stats.do	FigureA11_A.pdf
Figure A11 (panel B)	analyze_robust.do	FigureA11_B.pdf
Figure A12	descriptive_stats.do	FigureA12_A.pdf, FigureA12_B.pdf

Table 10. Appendix B Exhibits

Exhibit	Script	Output file(s)
Table B1	analyze_hetero.do	tableB1.csv
Figure B1 A–B	analyze_hetero.do	FigureB1_A.pdf, FigureB1_B.pdf
Table B2	analyze_hetero.do	tableB2.csv
Figure B2 A–B	analyze_hetero.do	FigureB2_A.pdf, FigureB2_B.pdf
Table B3	analyze_hetero.do	tableB3.csv
Figure B3 A–B	analyze_hetero.do	FigureB3_A.pdf, FigureB3_B.pdf

Table B4	analyze_hetero.do	tableB4.csv
Figure B4 A–B	analyze_hetero.do	FigureB4_A.pdf, FigureB4_B.pdf
Table B5	analyze_hetero.do	tableB5.csv
Figure B5 A–B	analyze_hetero.do	FigureB5_A.pdf, FigureB5_B.pdf
Table B6	analyze_main.do	tableB6.csv
Figure B6 A–H	analyze_main.do	FigureB6_A.pdf, ..., FigureB6_H.pdf
Table B7	analyze_hetero.do	tableB7.csv
Figure B7 A–B	analyze_hetero.do	FigureB7_A.pdf, FigureB7_B.pdf
Table B8	analyze_hetero.do	tableB8.csv
Figure B8 A–B	analyze_hetero.do	FigureB8_A.pdf, FigureB8_B.pdf
Table B9	analyze_hetero.do	tableB9.csv
Figure B9 A–B	analyze_hetero.do	FigureB9_A.pdf, FigureB9_B.pdf
Table B10	analyze_hetero.do	tableB10.csv
Figure B10 A–B	analyze_hetero.do	FigureB10_A.pdf, FigureB10_B.pdf
Table B11	analyze_hetero.do	tableB11.csv
Figure B11 A–B	analyze_hetero.do	FigureB11_A.pdf, FigureB11_B.pdf
Table B12	analyze_hetero.do	tableB12.csv
Figure B12 A–B	analyze_hetero.do	FigureB12_A.pdf, FigureB12_B.pdf