A Retrospective Analysis of the Acquisition of Target's Pharmacy Business by CVS Health: Labor Market Perspective

Enas Farag [*] Alaa Abdelfattah[†] Chris Compton[‡] Anna Stansbury [§]

Marshall Steinbaum [¶]

March 3, 2026

RESTAT-30863: ReadMe File for Data and Code Disclosure

[*]Department of Economics, University of Illinois Urbana-Champaign and Faculty of Economics and Political Science, Cairo University.

[†]Occidental College.

[‡]Lightcast.

[§]MIT Sloan School of Management.

[¶]University of Utah Department of Economics.

1 Introduction

This ReadMe file describes how to replicate the analysis and exhibits in "A Retrospective Analysis of the Acquisition of Target's Pharmacy Business by CVS Health: Labor Market Perspective." The main, proprietary dataset used is the 2024 Lightcast "US Postings" dataset. We also use an older vintage, dated 2022, to create a crosswalk between the 2018 and 2010 Standard Occupational Classifications, as described in Appendix C of the paper.

2 Obtaining the proprietary data

Since the paper uses proprietary data from Lightcast, researchers wanting to replicate or extend our results will have to acquire the raw data from the vendor. Either contact the company using their online interface, available at https://lightcast.io/contact-us, or email Chris Compton (chris.compton@lightcast.io). In addition, we are grateful to the authors of Schubert, Stansbury and Taska (2024) for sharing their occupational mobility data that use 16 million unique US workers' resumes collected by Lightcast. Access requests for these data should be directed to Anna Stansbury (amms@mit.edu).

3 Software details

Analysis was performed in Stata 18/MP running on Windows Server 2019 Standard edition. Hardware used for the last replication:

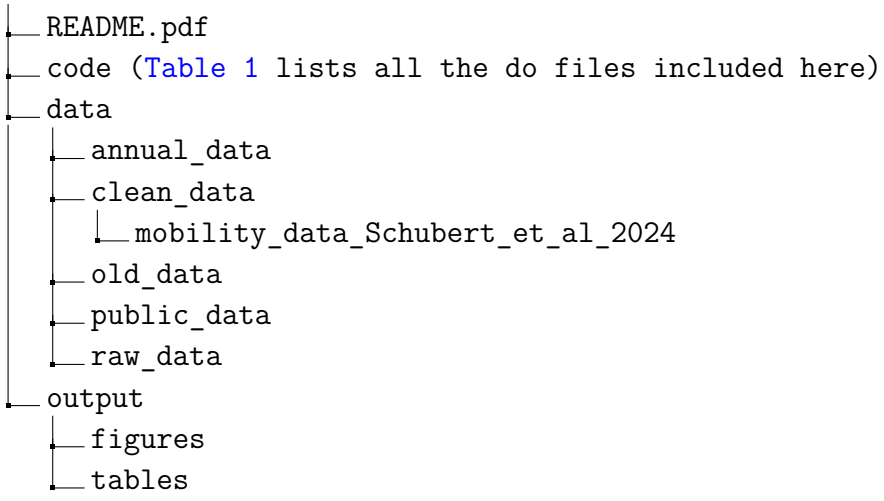
- Processor: Intel Xeon Gold 6254 CPU @ 3.09 GHz
- Memory: 256 GB RAM
- Computing environment: single local machine (no cluster or distributed computing)
- Storage requirement: 597 GB.

– This is the size of the replication package when the data folder has all the the proprietary datasets (raw and cleaned).

4 Folder Directory

Our replication package consists of three main folders: code, data, and output. Since we use proprietary data, our data folder is mostly empty —except for a sub-folder titled "public_data" which contains a few publicly available datasets from the Bureau of Labor Statistics (BLS) and the Quarterly Workforce Indicators (QWI) which we use in some supplementary analysis. We recommend for any replicator to use the same folder structure we use which is depicted in Figure 1 so that all the do files run smoothly.

Figure 1. Replication Package Directory Structure



5 Do Files Directory

Table 1 below explains the purpose of each STATA do file and the exhibits it generates. The master do-file "0.Master_do_file.do" allows the replicator to install necessary packages and set global directories to run all the do files smoothly without the need to change anything else. To

do so, update **line 38** in the Master do file located in the code folder.

Table 1. Description of Replication Scripts and Generated Exhibits

Script	**Description**	**Exhibits generated**
0.Master_do_file.do	Master do-file for the paper.	
Lightcast data cleaning code		
1.import_data_new.do	Import each .csv of raw data and label variables, output .dta	
2.create_time-sorted_files.do	Append hundreds of .dtas into annual datasets.	
3.merging_cz_variable.do	Import USDA's csv county-to-CZ crosswalk, outputs as .dta **and** merge cz variable to annual data.	
4.2010_2018_soc_crosswalk.do	Match 2018 SOC codes in LC data to 2010 SOC codes.	
5.non_missing_salary_data.do	Create a dataset of all vacancies with non-missing CZ and non-missing salary information for the period 2010–2024.	
6.tagging_pharmacy_chains.do	Tag employers of major pharmacy chains using the vacancies with *non-missing* salary information 2010–2024.	
7.2015_full_data.do	Construct the sample of 2015 job postings including the ones with *missing* salary that will be used later for tagging treated CZs and calculating Δ HHI.	
8.tagging_treated_CZs.do	Tag treated commuting zones.	
9.sample_construction _full_and_retail.do	Construct the main analysis samples: full sample (all industries) and retail sample (specific retail industries), as explained in subsection II.C.	
Paper-specific code		
10.occupational_leave _share_rank.do	Rank occupations (6-digit 2010 SOC codes) based on their leave share using Schubert, Stansbury and Taska (2024) mobility data.	

Continued on next page...

Table 1 continued from previous page		
Script	**Description**	**Exhibits generated**
11.occupational_salary_rank.do	Rank occupations (6-digit 2010 SOC codes) based on average annual earnings using 2015 BLS OES estimates.	
12.calculating_delta_CZ _2015.do	Compute predicted delta HHI for treated CZs (Equation II.1).	
13.ojoi_calculation.do	Compute outside job option index (OJOI) (Equation II.2).	
14a.did_regression _full_sample.do	Run simple DiD regression using full sample (subsection III.A).	Table 1 Table A2
14b.event_study _full_sample.do	Create event study plots for DiD regression using the full sample (subsection III.A).	Figure 1
15.did_regression _retail_sample.do	Run simple DiD regression using retail subsample (subsection III.B). Create event study plots for retail subsample.	Table 2 Figure 2
16a.retail_triple _difference_regression.do	Run retail DDD regression (subsection III.C).	Table 3
16b.retail_triple _difference_event_study.do	Create event study plots for retail DDD regression.	Figure 3
17a.treatment_intensity _HHI_regression.do	Run treatment intensity using predicted Δ HHI regression (subsection III.D).	Table 4
17b.treatment_intensity _HHI_event_study.do	Create event study plots for treatment intensity using predicted Δ HHI regression.	Figure 4
18a.treatment_intensity _OJOI_regression.do	Run treatment intensity using outside job option index (OJOI) regression (subsection III.E).	Table 5
18b.treatment_intensity _OJOI_event_study.do	Create event study plots for treatment intensity using OJOI regression.	Figure 5

Continued on next page...

Table 1 continued from previous page		
Script	**Description**	**Exhibits generated**
19.ddd_regression_by _salary_rank.do	Run retail DDD regression interacted with below- and above- median salary indicator to test for heterogeneous merger effects by occupational salary rank.	Table 6
20.ddd_regression_by _leave_share_rank.do	Run retail DDD regression interacted with below- and above-median leave share indicator to test for heterogeneous merger effects by occupations' degree of outward mobility.	Table 7
21.figure6_coefplots.do	Create coefficient plot visualizing the estimated merger effects across different research designs and by occupational heterogeneity.	Figure 6
A1.figure_A1.do	Create a plot of total vacancy count as a share of total employment for full sample and retail subsample.	Figure A1
A2.appendix_figures_tables.do	Create appendix figures and tables.	Figure A2 Figure A3 Figure A4 Figure A5 Table A1 Table A3
B.ddd_regression _MSA_FEs.do	Run retail DDD regression with MSA-by-YQ FEs. Create event study plots for retail DDD with MSA-by-YQ FEs.	Table B1 Figure B1

6 Data dictionary

The following table identifies the variables included in the data used for this paper, which was obtained in July 2024. We have observed that different vintages of Lightcast data may include different variable names for the same underlying variable, and more significantly, includes

additional fields as time goes on. The fields named in the following table are those from the vintage used for the present project.

Table 2. Lightcast data dictionary

Lightcast field	**Description**
id	Unique ID generated by Lightcast
employer	Employer name
soc	SOC code of the job, assigned using Lightcast occupation coding rules, based on the 2018 SOC delineations
naics4	4-digit NAICS code
county	5-digit code, representing the concatenation of the state and county FIPS codes
msa	Metropolitan Statistical Area code
salary_from	Minimum annual salary. If hourly rate is provided, this rate is converted to annual salary through multiplying the rate by 2080
salary_to	Maximum annual salary. If hourly rate is provided, this rate is converted to annual salary through multiplying the rate by 2080
salary	Mean of the salary range
date	Date the posting was spidered
cz	Commuting zone code using USDA, Economic Research Service's (ERS) 2000 commuting zones delineation