How Does Online Learning Affect Business Formation, Productivity, and Employment?

Octavio M. Aguilar

December 12, 2024

1 Repository Description and Initial Steps

I include the necessary data to replicate my main results. However, if you would also like to replicate my data cleaning processes, you will need to take a few extra steps. I do not include some of the raw files due to their large size. As a result, you will need to downloaded them on your computer and place them in the correct folder. Below, I detail what files live in each folder, how to extract the raw data, and how to replicate my main results. If there are any questions please email my personal: octaviomaguilar@outlook.com

1. The repository Replication-Online-Learning-Dynamism has 3 folders.
   (a) **Data**. In this folder you will find 12 subfolders:
      i. **atus:** This folder contains all the American Time Use Survey (ATUS) data. I do not include the raw ATUS data file due to its large size. Instead, I include "atus_trimmed.dta". This is a trimmed dataset. To pull the raw data please visit the IPUMS-ATUS website `https://www.atusdata.org/atus/` and extract the necessary variables.

      ii. **bartik:** This folder contains the shift-share (bartik) measures and clean datasets with the merged bartik measure.
         - bartik_naics2.dta: Bartik shift-share at the 2-digit industry level.
         - bartik_naics3.dta: Bartik shift-share at the 3-digit industry level.
         - bartik_naics4.dta: Bartik shift-share at the 4-digit industry level.
         - bartik_qwi_earn_fage_trimmed.dta: A clean dataset with the Bartik shift-share at the 4-digit industry level and earnings by firm age group.

- bartik_qwi_earn_fsize_trimmed.dta: A clean dataset with the Bartik shift-share at the 4-digit industry level and earnings by firm size group.
- bartik_state_naics4_fage_trimmed.dta: A clean dataset with the Bartik shift-share at the 4-digit industry level and employment share by firm age group.
- bartik_state_naics4_fsize_trimmed.dta: A clean dataset with the Bartik shift-share at the 4-digit industry level and employment share by firm size group.

iii. **bds:** This folder contains all the Business Dynamics Statistics (BDS) data. I do not include the raw BDS data files due to its large size. Instead, I include trimmed versions of the data. To pull the raw data please visit the BDS website: `https://www.census.gov/data/datasets/time-series/econ/bds/bds-datasets.html` You will need to extract the following datasets from the website:

- Sector by Firm Age Coarse
- Sector by Firm Age
- Sector by Firm Size
- 4-digit NAICS by Firm Age Coarse

Once you extract the raw data please place them into this BDS folder and save as a CSV.

iv. **bed:** This folder contains all the Business Employment Dynamics (BED) data. I do not include the raw BED data file due to its large size. Instead, I include a clean version of the data, bed_industry.dta. To pull the raw data please visit the BED website: `https://download.bls.gov/pub/time.series/bd/`. You will need to download the following: bd.data.0.Current It will download as a CURRENT file. This is fine. Once downloaded place it into this folder.

- bed_industry.dta: This data file is the clean BED data containing a panel of industries and their dynamism measures.
- bed_atus_educ.dta: This data file contains the online learning rate from 2003Q1 through 2022Q4 at the super sector level as well as the corresponding dynamism measures from the BED.
- bed_industry_year.dta: This data file contains the dynamism measures from the BED at the yearly frequency.

v. **bfs:** This folder contains data on business applications from the Business Formation Statistics (BFS).

vi. **bls:** This folder contains data on industry productivity from the BLS.
   A. blsip_naics3_regdata.dta:: This data file contains BLS industry productivity data and the bartik measure at the 3-digit industry level.
   B. blsip_naics4_regdata.dta: This data file contains BLS industry productivity data and the bartik measure at the 4-digit industry level.

vii. **ces:** This folder contains the CES reverse ratio at the 4-digit industry level. This is used when cleaning and calculating employment shares. For more information on how the CES reverse ratio is constructed and why it is used visit: `https://www.bls.gov/ces/naics/`

viii. **cps:** This folder contains the clean data cps_delta_educ.dta, which are the weighted 2-digit occupational shares of online learning in 2019 and 2021 as well as the changes in online learning between the two years. I do not include the raw CPS data files due to its large size. To download the raw data please visit the CPS-CIS website: `https://www.census.gov/data/datasets/time-series/demo/cps/cps-supp_cps-repwgt/cps-computer.html` Here you will download the 2019 and 2021 supplement as a CSV file. Once you download them please place them into this folder as a CSV.

ix. **crosswalks:** This folder contains all the crosswalks needed for the data cleaning process.

x. **oes:** This folder contains files used in the creation of the bartik variable. In order to create the bartik at the 3-digit and 4-digit industry level, you will need to download the raw data from the Occupational Employment and Wage Statistics. The file size is too big to keep in this folder. Please visit this url: `https://www.bls.gov/oes/tables.htm` and download the May 2019 national file. Convert it to a Stata dta file then store and name it here: "data/oes/nat4d_2019.dta". For the bartik at the 2-digit industry level, the file "natsector_M2019_dl.xlsx" is provided.

xi. **qcew:** This folder contains the file qcew_naics3_emp_all.dta. This contains total employment by 3-digit industry. This is used as a weight in the empirical specifications to account for differences in industry sizes.

xii. **qwi:** This folder should be empty. I do not include the raw QWI files due to its large size. You will need to download the

data yourself and place it into this directory. To download the data please visit the QWI LED website: `https://ledextract.ces.census.gov/qwi/all` You only need to download the data if you wish to process the code from start to finish. Otherwise, you already have the final datasets created from the QWI. For specific download instructions, please visit the QWI folder in this repository. It will give you step by step instructions.

(b) **Figures**. In this folder you will find the main figures from the text. There is a corresponding DO file for each figure. Please see program descriptions below.

(c) **Programs**. In this folder you will find the programs used to clean and analyze the data. In particular, there are two subfolders.

i. **Clean:** This folder contains the Stata programs to import and clean the data.

- S1_create_delta_learning.do: This program will create the "shift" in the shift-share bartik measure. Specifically, it will calculate the changes in online learning from 2019 to 2021 across 2-digit occupations.

- S2_create_bartik_naics3.do: This program will create the bartik measure at the 3-digit industry level by interacting (i) OES 3-digit industry-occupation employment shares with (i) the shift from S1.

- S3_create_bartik_naics4.do: This program will create the bartik measure at the 4-digit industry level by interacting (i) OES 4-digit industry-occupation employment shares with (i) the shift from S1.

- S4_create_bartik_naics2.do: This program will create the bartik measure at the 2-digit industry level by interacting (i) OES 2-digit industry-occupation employment shares with (i) the shift from S1.

- S5_create_blsip_naics4_regdata.do: This program will clean the raw BLS-IP data and merge in the bartik measure at the 4-digit industry level/

- S6_create_bartik_state_naics4_emp_fsize.do: This program will clean the QWI employment share by firm size data and merge in the bartik measure at the 4-digit industry level.

- S7_create_bartik_state_naics4_emp_fage.do: This program will clean the QWI employment share by firm age data and merge in the bartik measure at the 4-digit industry level.

- S8_import_qcew.do: This program will import and clean the QCEW data,

- S9_create_qcew_sector_naics3.do: This program will generate total employment across all 3-digit industries. This is used as weight in the empirical specifications to account for industry size.

- S10_create_bed_industry.do: This program will clean the raw BED data and create a panel of industries with business dynamism measures.

- S11_create_atus_educ_industry.do: This program will clean the raw ATUS data and calculate the online learning rate–as defined in the main text.

- S12_create_bed_atus_educ.do: This program will merge clean ATUS data on the online learning rate with the clean BED data. This will result in a panel of industries with information on online learning rates and business dynamism measures.

- S13_create_qwi_bartik_earn_fage.do: This program will clean the QWI earnings by firm age data and merge in the bartik measure at the 4-digit industry level.

- S14_create_qwi_bartik_earn_fsize.do: This program will clean the QWI earnings by firm size data and merge in the bartik measure at the 4-digit industry level.

ii. **Analysis:** This folder contains the Stata programs used to analyze the data. Mainly, it creates all figures and tables in the main results PDF. The name of the programs are self explanatory. The only programs I will note are:
   - figures_dev_empshare_fage.do: This program creates the event study figures that investigate the deviation in employment share by firm age groups. It outputs a combination of figures 4B and 5C and 5D
   - figures_dev_empshare_fsize.do This program creates the event study figures that investigate the deviation in employment

share by firm size groups. It outputs a combination of figures 4A and 5A, 5B and 5E.

2 Replication Instructions

Make sure that the repository is downloaded as provided and all files are in the correct folders. The programs are all automated and should be easily executable. The only thing you must do is change the global home directory to match your computer. You only need to change this one global. All other file paths build off this one. For example, you only need to change: global home = "/mq/scratch/m1oma00/oma_projects/replication_online_productivity_oma_final" to global home = "/this/is/where/I_put_the/repository".

1. After changing the home directory, for the quickest and easiest replication, one can run the Master.do program. If you DO NOT have all the raw data files you will only be allowed to run the second block in the master program, that is, the analysis. To run the master from start to finish (cleaning to analysis) you will need to download all raw data files as described above.

2. If you have all the raw data and wish to run each program one by one, start with the cleaning programs. Run them in order from S1 through S14. You may then run any analysis program you wish.

3. Alternatively, if you do not want to download the raw data, after downloading the repository you can skip the cleaning programs and run any analysis program you wish.