Restaurant employment, minimum wages, and border discontinuities – documentation for replication files

Summary

This folder contains code and datasets to replicate the exhibits in Dube, Reich, Bhatt and Sosinkiy (2025) – Restaurant employment, minimum wages, and border discontinuities.

The *data/raw* folder contains all inputs you will need, while the *data/clean* folder has datasets saved after cleaning. Tables and figures produced during analysis are in the *out* folder.

There is code for cleaning raw data, in *code/cleaning,* and code for analysis in *code/analysis*. The cleaning and analysis files are named according to the order in which they should be run if you want to replicate the whole analysis from scratch. However, it is possible to run any individual files as well (all the data that any files will need has been saved, even if it is originally formed in an upstream file).

To run any files, you need to first run 00_path_master.do. This sets global macros for filepaths. You can run files individually, or through the 01_run_all.do file.

The table below shows the analysis files that are used to create different exhibits in the paper. All of these do files draw on clean data saved using various cleaning do files. Table B1 – which relates to sample descriptions – is the only table formed in a cleaning file (06_prep_samples_qcew)

Exhibit	Do files used
Figure 1	01_twfe_cbp (saves results), 02_twfe_qcew (outputs final figure)
Figure 2	06_event_study_qcew
Figure 3	06_event_study_qcew
Table 1	01_twfe_cbp
Table 2	02_twfe_qcew
Table 3	06_event_study_qcew
Figure A1	07_ever_treated_vs_never_treated_qcew
Figure A2	08_maps
Table A1	01_twfe_cbp
Figure A3	01_twfe_cbp (saves results), 02_twfe_qcew (outputs final figure)
Figure A4	01_twfe_cbp (saves results), 02_twfe_qcew (outputs final figure)
Table A2	03_twfe_qcew_quarterly
Table A3	03_twfe_qcew_quarterly
Table C1	09_twfe_binary_sim
Figure C1	09_twfe_binary_sim
Figure D1	06_event_study_qcew
Table D1	06_event_study_qcew
Table D2	06_event_study_qcew
Table D3	06_event_study_qcew
Table D4	06_event_study_qcew
Table D5	06_event_study_qcew
Table D6	06_event_study_qcew

Restaurant employment, minimum wages, and border discontinuities – documentation for replication files

Table D7	06_event_study_qcew
Figure E1	10_actualmw_sim

Details of files

Data folder:

Raw – QCEW

- Annual QCEW files according to NAICS classification (1990-2019)
- Quarterly QCEW files according to NAICS classification (1990-2019)
- Appended versions of annual and quarterly files that only keep data relevant to us

Raw – JNR-CBP

This folder contains data taken from JNR's replication files, which we use for all CBP exhibits

Raw – others

- 16+ population by state – from CPS
- Dataset denoting start of MW indexation in states with indexation
- Annual state level MW data from Vaghul and Zipperer (2022)
- Quarterly state level MW data from Vaghul and Zipperer (2022)
- Consumer price index
- State abbreviations
- SEER county level population data
- List of events identified in Cengiz, Dube, Lindner, Zipperer (2019)
- Map coordinates of states, counties, commuting zones – downloaded from https://michaelstepner.com/maptile/geographies/

Clean data

These folders contain datasets outputted through the cleaning do files

Output folder

This folder contains final tables and figures. It also contains a sub-folder for intermediate results – these are useful when we are combining datasets for a figure/table.

Code folder

Cleaning

Restaurant employment, minimum wages, and border discontinuities – documentation for replication files

- 01_seer_population_data: Saves 16+ population for counties from raw SEER dataset (for QCEW)
- 02_clean_qcew_annual_county: Appends QCEW annual county level data and merges it with other datasets
- 03_clean_qcew_quarterly_county: Appends QCEW quarterly county level data and merges it with other datasets
- 04_ak_de_ri_90s: Corrects for missing QCEW state-level data for Alaska, Delaware and Rhode Island in the 1990s (as described below)
- 05_clean_qcew_annual_state: Appends QCEW annual state level data and merges it with other datasets
- 06_prep_samples_qcew: Creates CZ-state, border county pairs, multi-state commuting zone pairs, and state level samples for analysis from QCEW annual data
- 07_prep_samples_quarterly_qcew: Creates CZ-state, border county pairs, multi-state commuting zone pairs samples for analysis from QCEW quarterly data
- 08_prep_samples_cbp: Creates CZ-state, border county pairs, multi-state commuting zone pairs, and state level samples for analysis from CBP data
-

Analysis:

- 01_twfe_cbp: Conducts TWFE analysis using CBP data including replicating JNR's results
- 02_twfe_qcew: Conducts TWFE analysis using annual QCEW data
- 03_twfe_qcew_quarterly: Conducts TWFE analysis using quarterly QCEW data (for some robustness tables in the appendix)
- 04_create_events_cbp: Constructs events for MSCZ dataset using CBP data to be able to use for analysis of pre-trends in 06_event_study_qcew
- 05_cdlz_events: Conducts event study analysis using events from Cengiz, Dube, Lindner, Zipperer (2019) for a robustness check
- 06_event_study_qcew: Conducts all event study analysis, including checking for pre-trends, and robustness
- 07_ever_treated_vs_never_treated_qcew: Generates Figure A1 from state-level data (that includes Alaska and Hawaii, unlike all other analysis, following Dube and Lindner (2024)).
- 08_maps: Generates the maps in Figure A2
- 09_twfe_binary_sim: Runs simulation reported in Appendix C
- 10_actualmw_sim: Runs simulation reported in Appendix E

Details of issues in cleaning QCEW county-level data

This section describes some cleaning made in the QCEW cleaning do files. Mostly, these pertain to issues with merging counties across datasets. We use different sources for our employment and earnings numbers, population, and commuting zones. In this document, we treat the QCEW data

constructed from annual QCEW files as the baseline and describe the issues and resolutions in merging different datasets with it. Overall, two sources were crucial for resolving most issues –

1) https://www.ddorn.net/data/FIPS_County_Code_Changes.pdf - a document by David Dorn summarizing the change in county definitions that have taken place since the 1980s
2) https://seer.cancer.gov/seerstat/variables/countyattribs/ruralurban.html, where we got our population data from. They describe some changes to countyfips they make in absence of mortality and population data

Merging commuting zones

We merge commuting zone data from JNR's replication files.

Here, we initially found three merge == 2 counties (i.e they are in the JNR dataset but not in our dataset). These are 51780 in Virginia, 30113 in Montana, and 12025 in Florida. Apart from this, there is one merge == 1 county. However, if you look at the Dorn document linked above – all these three counties figure in the 1990s changes section – all three of them were merged into larger counties. We checked and we do have exact merges for all these larger counties, so no change required.

We also initially found a merge == 1 county (i.e it exists in the QCEW dataset, but not in the JNR dataset). This was FIPS 46102 or "Oglala Lakota County, South Dakota". Again, as seen in the Dorn document, FIPS 46113 or Shannon County was renamed as this in 2015. This tracks with our QCEW dataset, where we had data for 46113 till 2015, and data for 46102 from 2015 onwards. **We just changed the FIPS in the QCEW data from 46102 to 46113 for all years (as the JNR data has 46113). For some reason, 2015 data was present under both FIPS with very slight differences. We kept the one that was listed under 46102 (the new FIPS) for 2015.**

Merging population data

We use population data from seer.cancer.gov (the do file seer_population_data.do processes this to keep only working age population).

Colorado

Firstly, we get some issues for Colorado. We get four counties (FIPS – 8911, 8912, 8913, and 8914) that are merge == 2 (exist in pop data, but not in QCEW data) for years 1990-2001. We also get five counties (FIPS – 8001, 8013, 8059, 8123, 8014) that are merge == 1 (are in QCEW but not in pop data) for the same range of years. Looking at the SEER document linked above solves the puzzle. This is what they say:

Broomfield, Colorado (08014) became a county in November 2001. Mortality data became available for Broomfield in 2003. Due to the fact that this county was created from parts of four other counties (Adams, Boulder, Jefferson, Weld), new counties were created to address the fact that the boundaries have changed over time.

Restaurant employment, minimum wages, and border discontinuities – documentation for replication files

- *Adams county is represented by a FIPS code of (08911) for data from 1969-2002 and the Rural-Urban Continuum code for (08911) matches (08001) for all years.*
- *Boulder county is represented by a FIPS code of (08912) for data from 1969-2002 and the Rural-Urban Continuum code for (08912) matches (08013) for all years.*
- *Jefferson county is represented by a FIPS code of (08913) for data from 1969-2002 and the Rural-Urban Continuum code for (08913) matches (08059) for all years.*
- *Weld county is represented by a FIPS code of (08914) for data from 1969-2002 and the Rural-Urban Continuum code for (08914) matches (08123) for all years.*

Thus, what we do is - change FIPS 8911, 8912, 8913, and 8914 to the FIPS for Adams, Boulder, Jefferson and Weld counties (8001, 8013, 8059, and 8123) in the population data. This enables us to get a complete match for these counties. We still couldn't get a match for Broomfield county.

Arizona

For Arizona, we have a merge == 2 county from 1990-1993 with FIPS 4910. We also have two merge == 1 counties in the same year range (FIPS 4012 and 4027). These two counties merge in for the rest of the years (that is we have population data for them in all years apart from 1990-1993). Again, the SEER document and the Dorn document gives some context:

Arizona, 1983: La Paz county (FIPS 4012) is created out of parts of Yuma county (FIPS 4027). – Dorn doc

Since county data is not available in our data until 1994, we needed to create a code for the grouped county La Paz/Yuma (04910) 1969-1993 – SEER doc

This just means that the 4910 merge == 2 that we have for 1990-1993 is combining data for La Paz and Yuma counties. Fortunately, we have population data for these counties from 1994 onwards. **So what we do is this – we divide the 1990-1993 combined population of the two counties in accordance with the ratio of their populations in 1994 (the first year we have separate population).**

Virginia

Restaurant employment, minimum wages, and border discontinuities – documentation for replication files

Similar to Arizona, we have three merge == 1 counties – 51019, 51515, and 51560 (Bedford County, Bedford City, Clifton Forge City) – for all years. We also have one merge == 2 FIPS code for all years. This is what the two linked documents tell us:

Virginia, 2001: The independent city of Clifton Forge (FIPS 51560) merges into Alleghany county (FIPS 51005) – Dorn doc

Virginia, 2013: The independent city of Bedford (FIPS 51515) merges into Bedford County (FIPS 51019) – Dorn doc

Since individual county data is not available in our data until 1980, we need to create a code for the grouped county, Bedford City and County (51917) 1969-1979. – SEER doc

The Clifton Forge case is fine, because we get a full match for Alleghany county (and Clifton Forge anyway had data only for one year).

The Bedford County and Bedford City case seem like Arizona above, but there are two issues. Firstly, even though the SEER doc says they use the combined code 51917 only till 1979, the data has only 51917 until 2019. There is no 51515 or 51019 in the SEER data (hence, Bedford city and Bedford county are merge == 1 for all years). This means that unlike La Paz and Yuma, we have no year in which we have population data for both counties separately to create a ration. **Hence, we have to drop these counties.**

Details of issues in cleaning QCEW state-level data

While most of our analysis uses county-level data (apart from Figure A1, which is adapted from Dube and Lindner (2024)), we still clean state level QCEW data, as it is a useful baseline for constructing minimum wage increase events.

The state-level data has one issue - three states---Alaska, Delaware, and Rhode Island---have missing or incomplete data for the restaurant sector (NAICS code 722) in the 1990s.[1] For these states, we impute data for the 3-digit restaurant sector using available data from 4-digit sub-sectors. In the NAICS classification system, there are five 4-digit sub-sectors within the broadly defined restaurant sector (the description for NAICS code 722 is ``Food services and drinking places''). These 5 sub-sectors are 7221 (Full-service restaurants), 7222 (Limited-service eating places), 7223 (Special food services), 7224 (Drinking places), and 7225 (Restaurants and other drinking places). Out of these, data on 7221 and 7222 is consistently available through the 1990s for all three of Alaska, Delaware, and Rhode Island. The other three sub-sectors are not available in one or more of the three. Thus, we use only 7221 and 7222 for our imputation.

[1] We do not need Alaska for analysis in the paper as we drop both Alaska and Hawaii, but we do need it to reproduce Figure A1 similar to Dube and Lindner (2024).

Clearly, just adding these two sub-sectors is not enough as they leave out some components of 722, which would create non-comparability of levels around 2000. To solve this, we take data for 2001 (which is the first year with full data), and take the ratio of employment under the 3-digit code 722 to the sum of employment under the 4-digit codes 7221 and 7222. We then multiply the 7221 + 7222 employment numbers from 1990-2000 by this factor. This gives us an estimate for employment for the full 3-digit sector in the 1990s for these states. All this analysis is done in cleaning do file "04_ak_de_ri_90s"

References

Cengiz, Doruk, Arindrajit Dube, Attila Lindner, and Ben Zipperer. "The effect of minimum wages on low-wage jobs." *The Quarterly Journal of Economics* 134, no. 3 (2019): 1405-1454.

Dube, Arindrajit, and Attila Lindner. "Minimum wages in the 21st century." *Handbook of Labor Economics* 5 (2024): 261-383.

Vaghul, Kavya and Ben Zipperer (2022) "Historical State and Sub-state Minimum Wages," Version 1.4.0, https://github.com/benzipperer/historicalminwage/releases/tag/v1.4.0.