## Overview of files and folders in .\prices

Files in this folder are used to create childcare cost and household goods/service input prices.

- Childcare costs are obtained from Child Care Aware (CCA) of America's Cost of Child Care Reports covering years 2006-2018. After some cleaning, all missing values and earlier years are imputed based on regressions on some or all of the following: state dummies, year time trends, state-specific time trends, and average wages for childcare workers (and preschool and kindergarten teachers) by state and year (from monthly CPS). See "Additional notes on childcare costs" below for further details.

- Measures for "all goods and services", "goods", and "services" at the state-year level are obtained by combining annual US average Consumer Price Index (CPI-U) measures with relative price measures by state-year from the Bureau of Economic Analysis Regional Price Parities (RPP). See "Additional notes on CPI and RPP" below for further details.

- Main output files are p_extendedcosts.dta and p_extendedcosts.csv, which contain several different childcare cost measures (actual and imputed), prices for goods and services—all for 1992-2019

All folder references below begin with root folder: .\FINAL_Replication_CLMP_JPE

**Code**

- *main.do*
  - sets folder structure and executes all other .do files in this folder.
  - Must set $root folder.

- *Read_cps_00008.do*
  - reads in raw CPS ascii data file: .\data\data_CPS\cps_00008.dat
  - formats and labels variables

- *Clean_CPS.do*
  - cleans CPS data, selects relevant sample of childcare (and related) workers
  - creates hourly wage measures

- *CPS_Wage_v2.do*
  - calculates average wages for childcare workers by state and year
  - saves these average wages by state and year in
    - .\code\data_cleaning\prices\temp_data\wage_CC_data.dta/xls – contains average hourly wages by state-year for childcare workers
    - .\code\data_cleaning\prices\temp_data\wage_CCPSKT_data.dta/xls – contains average hourly wages by state-year for childcare workers + preschool and kindergarten teachers

- *ReadCosts.do*
  - imports original CCA childcare cost data from .\data\ChildCareAware\ALLCosts_forSTATA.xlsx
  - saves .\code\data_cleaning\prices\temp_data\childcost_raw.dta

- *ReadStateCPI.do*
  - imports .\data\CPI_RPP\Relative_Prices_by_State2_FINAL_forSTATA.xlsx
  - saves .\code\data_cleaning\prices\temp_data\stateCPI.dta

- *ReadWage.do*
  - reads .\code\data_cleaning\prices\temp_data\wage_CC_data.dta/xls and .\code\data_cleaning\prices\temp_data\wage_CCPSKT_data.dta/xls
  - merges and saves .\code\data_cleaning\prices\temp_data\wage_raw.dta

- *MergeChildWageCPI.do*
  - merges temporary data sets created above containing different input costs and childcare worker average wages by state-year
  - saves these in temporary data set .\code\data_cleaning\prices\temp_data\extendedcosts.dta

- *Predicted_series.do*
  - reads .\code\data_cleaning\prices\temp_data\extendedcosts.dta
  - runs prediction regressions to impute missing childcare costs
  - saves childcare, goods, and services input prices at the state-year level in data set used for main empirical analysis: .\data\data_derived\p_extedendcosts.dta/csv
  - see "Summary of p_extendedcost.dta" below for more details on these data

**Other folders**

- .\code\data_cleaning\prices\logs – Contains log files produced by .do files described above

- .\code\data_cleaning\prices\temp_data – Used as a repository for temporary data files when running Stata .do programs

**Related data folders**

- .\data\data_ChildCareAware – Contains data on childcare costs from CCA
    - "CCA Long term childcare cost data (original).xlsx" – original file sent from CCA
    - AllCosts_FINAL.xlsx – original CCA data with noted cleaning efforts and comments
    - AllCosts_forSTATA.xlsx – cleaned cost data from CCA, used to import into Stata
    - Stateinfo.dta/xlsx – file with state codes used in matching data sets

- .\data\data_CPI_RPP
    - "Relative Prices by State2.xlsx" – contains CPI-U for all US 1983-2019 and RPP for all states from 2008-2017. Calculates state-year level prices for goods and services from 1983-2019 using year-over-year changes in regional CPI-U measures prior to 1983 and after 2017. See notes below (Additional notes on CPI and RPP) and Excel file for details.
    - Relative_prices_by_State2_FINAL_forSTATA.xlsx: contains final measures for "all goods & services", "goods", and "services" prices by state-year (taken from "all (cpi) REGIONS", "goods (commodities) REGIONS", and "serv o (serv less rent) REGIONS" worksheets of "Relative Prices by State2.xlsx"

- .\data\data_CPS
    - cps_00008.dat – raw monthly CPS data (1990-2019) used to calculate wages of childcare workers
    - data downloaded March 14, 2023, from IPUMS CPS website
    - see IPUMS_CPS_Extract_info.pdf for details on the IPUMS data extract request.
    - "CPS data notes.docx" contains detailed information on extracted variables.

## Summary of "p_extendedcosts.dta"

This file contains data at the state-year level for:

- The four main child cost measures: Infant Center Childcare, Infant Family Childcare, 4 year-old Center Childcare, and 4 year-old Family Childcare.
- Each of the childcare cost measures were separately imputed for missing years using predicted values from OLS regressions based on different sets of regressors. For some predictions, average wages for childcare workers in the CPS were included as regressors.
- Prices (with predictions replacing missing values) are saved as p_"var"_"case"_CPS_"wage measure" where
    - "var":
        - icent – Infant Center Childcare
        - ifcc – Infant Family Childcare
        - yocent – 4 year-old Center Childcare
        - yofcc – 4 year-old Family Childcare
    - "case":
        - a - State dummies only
        - b - State dummies and linear trends
        - b1 - State dummies and linear trends + interaction
        - c - Childcare worker wages only
        - d - Childcare worker wages and state dummies
        - e - Childcare worker wages, state dummies and linear trends
        - e1 - Childcare worker wages, state dummies and linear trends + interaction
    - "wage measure"
        - cpkt - Childcare worker wages measures from occupations "childcare workers" and "preschool and kindergarten teachers"
        - cc - Only "childcare workers" occupation

- CPI price index (2000=1) for: All items (all), Goods (goods), and Other Services (servo) by state and year.

- The childcare worker hourly wage measures used for the childcare price projections: cps_cc, cps_cpkt

# Additional notes on childcare costs

## CPS wage measures

- creates average state-year measures of hourly wages for following occupations (2010 occupational classification): (i) childcare workers and (ii) child care workers + preschool and kindergarten teachers
- uses data for individuals ages 18+ from 1992-2019 monthly CPS (IPUMS)
- wages taken from workers paid by the hour and those paid weekly divided by usual hours per week
- wages are trimmed by dropping top and bottom 1% of wage observations by year
- average state-year wages are calculated using CPS weights (earnwt)

## Predicted childcare cost series

- Child care costs for infant center care, infant family care, 4-year old center care, and 4-year old family care are obtained from Child Care Aware of America's Cost of Child Care Reports covering years 2006-2018
    - Several values from reports are dropped if they were imputed based on previous survey years or taken from different sources or subsets of locations (e.g. SD reports values for urban areas only in 2009, while California 2009 reports are from a separate survey)
- child care costs are regressed on some or all of the following: state dummies, year time trends, state-specific time trends, wages for childcare workers (and preschool and kindergarten teachers)
- Based on these estimates (and data on state-year average wages), childcare costs are predicted for years prior to the observed period.
- Predicted costs series are very highly correlated (corr>0.99) whenever we include state and time dummies in prediction regression
- Predicted cost series
    - has 1,412 year-state obs. with CPS measures using child care workers only (1992-2018)
    - has 1,428 year-state obs. with CPS measures using child care workers, kindergarten and preschool teachers (1992-2019)

## Additional notes on CPI and RPP data

The measures for the "goods" and "services" price series at the state level, are based on the Regional Price Parities by State (RPP) from the BEA, and the Consumer Price Index (CPI-U) from the BLS. The RPPs measure price level differences relative to the US average by state, and they are available from 2008-2017. These indices are divided in: All items; Goods; Services: Rent; and Services: Other.

To build our "goods" measure by state, we take the US average of the CPI for "commodities" and multiply it by each state's "Goods" RPP. This procedure allows us to have price measures by state for the period 2008-2017. To project back, we take the regional CPI for "commodities" and use the year-over-year change on this index for each state in that given region (Northeast, Midwest, South and West), starting from 2008. To project forward, we do a similar procedure starting in 2017. For our "services" measure, we use the "Services: Other" component from the RPPs, and the "Services less rent of shelter" index from the CPI.

All these prices are year averages. The base year for all the CPI series that we use is 2000.

What is included in these large categories?

These categories are special aggregate indices created by the BLS. From the table below, we are using the "Commodities" and the "Services less rent of shelter". The choice was based, in part, on the availability of the corresponding regional indices.

**Relative importance of components in the CPI: U.S. city average, December 2019**

| Special aggregate indexes | |
|---|---|
| All items | 100.000 |
| Commodities | 36.500 |
| Commodities less food and beverages | 22.266 |
| Nondurables less food and beverages | 12.580 |
| Nondurables less food, beverages, and apparel | 9.721 |
| Durables | 9.686 |
| Services | 63.500 |
| Rent of shelter | 33.204 |
| Transportation services | 5.876 |
| Other services | 11.596 |
| All items less food | 86.721 |
| All items less shelter | 66.429 |
| All items less medical care | 91.124 |
| Commodities less food | 23.222 |
| Nondurables less food | 13.535 |
| Nondurables less food and apparel | 10.676 |
| Nondurables | 26.814 |
| Apparel less footwear | 2.205 |
| Services less rent of shelter | 30.295 |
| Services less medical care services | 56.334 |
| Energy | 7.430 |
| All items less energy | 92.570 |
| All items less food and energy | 79.291 |
| Commodities less food and energy commodities | 19.078 |
| Energy commodities | 4.144 |
| Services less energy services | 60.213 |
| Domestically produced farm food | 5.992 |
| Utilities and public transportation | 9.228 |