In What Contexts do Households Maintain their Advantage? Examining Widening Household Expenditures on Higher Education across the Income Distribution

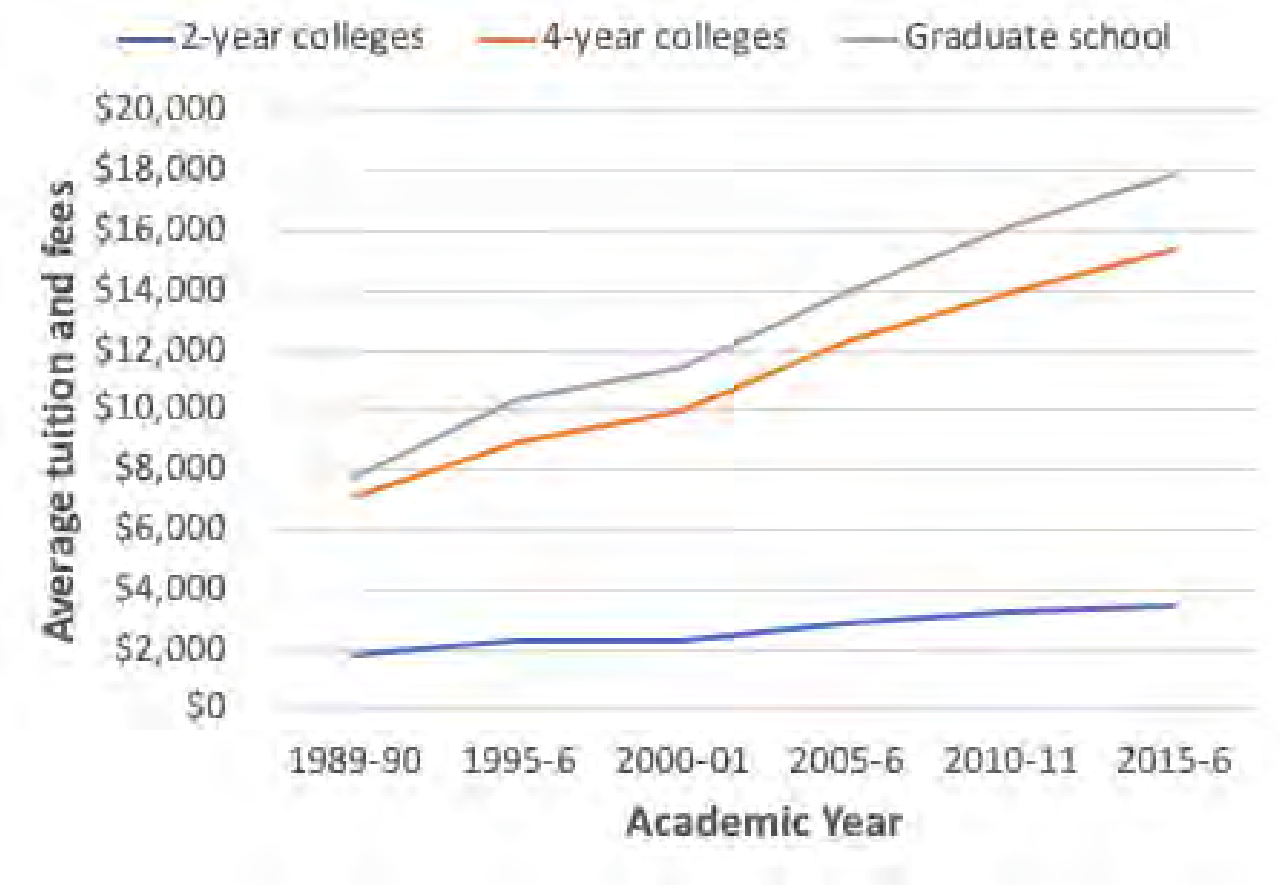
Kennan Cepa

University of Pennsylvania

Consumer Expenditure Survey Microdata Users' Workshop

July 17th, 2019

COLLEGE COSTS OVER TIME



IPEDS data from NCES (2016) summary of undergraduate and graduate school tuition and fees in constant 2015 dollars

COLLEGE COSTS ≠ FAMILY EXPENDITURES ON HIGHER EDUCATION

- IPEDS = institutional-level data about college tuition, enrollment, institution type, and other characteristics
  - Can be aggregated to state-level because every college that receives federal financial aid is compelled to respond to the survey
- CE = household-level data about higher education spending and enrollment
  - Includes a state identifier in the FMLI file

IPEDS & CE, TOGETHER

- If you want to:
  - Compare average college costs in a state to an individual household's spending on higher education or
  - Look at the postsecondary context in which families spend on college
- Then can merge CE (for household spending on higher education) with IPEDS (for information about college cost and postsecondary context)

FOCUSING ON CE HALF…

- Creating a sample using the interview survey files
- Getting the data ready:
  - Postsecondary expenditures
    - CE specifically asks families to include expenditures for children attending college who live away from home
  - State
  - Merge CE data together
  - Merge with IPEDS data
- Variable wish-list and questions
- Summary

CREATING A SAMPLE

- Households that have a child or grandchild 17+ years old currently enrolled in postsecondary institution
  - Requires information on:
    - Relationship of household members
    - Age
    - Enrollment information

CREATING A SAMPLE (CONT.)

- MEMI files come separately for each year/quarter and are organized as unique household members within unique households
  - **NEWID** = household identifier
  - **CU_CODE** = relationship of individuals within household
    - 3 & 4 are the codes for children & grandchildren
  - **IN_COLL** = college/university enrollment information for each family member
    - 1 & 2 are the codes for full-time and part-time enrollment
  - **AGE** = age of each family member
    - Restrict sample to those who are 17+

CREATING A SAMPLE (CONT.)

- Aggregate the MEMI file to unique year-household level

```
use "/Volumes/KENNAN CEPA/Raw Data/CE/intrvw17/intrvw17/memi171x.dta", clear

    gen coll_stud = 1 if (in_coll=="1" | in_coll=="2") & (cu_code=="3" | cu_code=="4") & age>=17

    bys newid: egen max_coll_stud = max(coll_stud)

    keep newid max_coll_stud

    ren max_coll_stud coll_stud

    duplicates drop
```

GETTING THE DATA READY

- Step 1: Clean the higher education expenditure variables
  - Use the EDA section of the EXPN file, which is organized as unique educational expenses per month within unique households every year
  - Restrict the sample to those with college-related expenses
  - Combine postsecondary expenses into an annual variable
  - Adjust expenses into constant dollars using the Consumer Price Index Research Series
  - Append all quarters and years together to make analytical sample

GETTING THE DATA READY (CONT.)

- Step 1: Clean the higher education expenditure variables
  - **EDSCHL_A**: Expenses are associated with a certain type of institution
    - 1 & 5 are college/university & vocational/technical schools
  - **EDMONTHA**: The month of the expense
    - 13 = monthly expense
  - **EDUC_AY**: Item code
    - 335 = total costs
    - 300 = tuition
    - 350 = books
    - 310 = housing
    - 320 = food
    - 370 = books and tuition
    - 380 = other costs
  - **JEDUCNET** = expense amount

GETTING THE DATA READY (CONT.)

- Step 1: Clean the higher education expenditure variables

```
use "/Volumes/KENNAN CEPA/Raw Data/CE/intrvw17/expn17/eda17.dta", clear

    keep if edschl_a=="1" | edschl_a=="5"
    //Only keeps obs that are a college/university or technical/vocational school expense

*Flags different types of college costs
    gen coll_all_cost = jeducnet if educ_ay=="335" & coll==1
    gen coll_tuition_cost = jeducnet if educ_ay=="300" & coll==1
    gen coll_books_cost = jeducnet if educ_ay=="350" & coll==1
    gen coll_housing_cost = jeducnet if educ_ay=="310" & coll==1
    gen coll_food_cost = jeducnet if educ_ay=="320" & coll==1
    gen coll_bookstuition_cost = jeducnet if educ_ay=="370" & coll==1
    gen coll_other_cost = jeducnet if educ_ay=="380" & coll==1

    foreach y in all tuition books housing food bookstuition other {
        replace coll_`y'_cost = coll_`y'_cost*12 if edmontha=="13"
    }
    //If the college cost takes place every month, multiply by 12 to get the total
    //amount spent in a given year

    foreach y in tuition books food housing {
        foreach x in 01 02 03 04 05 06 07 08 09 10 11 12 13 {
            gen `y'_cost_`x'17 = coll_`y'_cost if edmontha=="`x'"
        }
    }
    //Create a variable that holds the specific expense for each month
```

GETTING THE DATA READY (CONT.)

- Step 1: Clean the higher education expenditure variables

```
foreach y in tuition books food housing {
    foreach x in 01 02 03 04 05 06 07 08 09 10 11 12 13 {
        bys newid: egen max_`y'_`x'17 = max(`y'_cost_`x'17)
    }
}
//Fill so that each observation (which is unique expense in every household) holds the monthly
//expense for each broad category

foreach y in tuition books food housing {
    egen `y'_sum = rsum(max_`y'_0117 max_`y'_0217 max_`y'_0317 max_`y'_0417 max_`y'_0517 max_`y'_06
}
//This sums together expense in each month for each month
//Ie. All tuition spending for every month of the year is the variable tuition_sum
//////All book spending for every month of the year is the variable books_sum, etc.

keep newid *_sum
//Just keep the household ids & total variables for each type of expense in the year

duplicates drop
//Gets data to unique household level
```

GETTING THE DATA READY (CONT.)

- Step 1: Clean the higher education expenditure variables

	newid	tuition_sum	books_sum	food_sum	housing_sum
1	03386894	31	.	.	.
2	03386974	21876	.	.	.
3	03387584	1058	140	.	.
4	03387744	.	240	.	.
5	03388714	196	220	.	.
6	03390274	2196	.	.	.
7	03390434	.	10	.	.
8	03390504	400	.	.	.
9	03391024	.	300	.	.
10	03391294	24000	.	.	.
11	03391584	2852	148	.	.
12	03392124	475	.	.	.
13	03392874	21000	.	.	.

GETTING THE DATA READY (CONT.)

- Step 1: Clean the higher education expenditure variables
  - Some questions:
    - Starting in 2001, in the EXPN files, there are also questions about contributions for college students
      - Are those in addition to the costs collected in the EXPN files? Or are they the same as the "Other college expenses"? Should they be added to total from EDA section?
    - When analyzing parental expenditures on younger children, the literature sometimes uses a per-child measure of household spending.
      - Is this something that should be considered at the postsecondary level?

GETTING THE DATA READY (CONT.)

- Step 2: Clean the state variable
  - Use the FMLI file which includes the variables for state and household ID
    - **CU_CODE**: Household ID
    - **STATE**: State of residence
      - States are coded numerically (in alphabetical order) with some missing states because of small sample size
  - Check that file is at the unique household level

GETTING THE DATA READY (CONT.)

- Step 3: Merge all CE files together
  - MEMI files are now at unique household-quarter level, FMLI and EXPN are at the unique household-year level (because we coded them to be this way)
  - One issue facing CE users is how to merge these files together. Two methods used in the literature:
    - Treat each household-quarter as independent and interpret as quarterly spending (Schneider et al. 2018)
    - Annualize household expenditures by:
      - Averaging household expenditures and multiplying by 4 (Kornrich & Furstenberg 2013)
      - Multiplying household expenditures by 4 (Lunn & Kornrich 2018)
    - Does CE have a recommendation for users?

GETTING THE DATA READY (CONT.)

- Step 4: Merge CE with IPEDS
  - Download IPEDS files of interest, making sure to include the state variable
    - IPEDS's state variable is alphabetized so need to create a crosswalk to merge with numerical state variable in CE

IPEDS variable

stabbr	state
AK	02
AL	01
AR	05
AZ	04
CA	06
CO	08
CT	09
DC	11

CE variable

ADDITIONAL VARIABLES OF INTEREST

- Share of Spending on Higher Education to understand the financial burden of college for families
  - Households sometimes report higher education spending but have no household expenditures in the total expenditures variable
    - Why is this the case?
    - What is the best way to handle these sorts of discrepancies in the data?

ADDITIONAL VARIABLES OF INTEREST

- Starting in 2013, CE included a question about student loans
  - STUDFINX, STUDNTBX, STUDNTB, STDTYRB, STDTYRX, STDTYRBX in FMLI files
  - This reflects:
    - The interest paid on principal, late charges, and interest for student debt
    - Amount owed currently and a year ago
  - Would be helpful to have information about:
    - Loans taken out to pay for current enrollment (is that included in higher education expenditures in some way?)
      - Without it, we are underreporting household spending on college, especially for middle- and low-income families who are more likely to rely on debt
    - Any information on original date of loan borrowing or total amount borrowed
      - This could be helpful in understanding the *duration* of household spending on higher education

ADDITIONAL VARIABLES OF INTEREST

- Current higher education expenditure variables cannot discern between postsecondary level (two-year community college, four-year college, or graduate degree)
  - Would be a helpful update to have, especially to distinguish graduate school from undergraduate education

SUMMARY

- CE is a critical and underutilized tool for understanding the context of household spending on higher education
  - Lingering questions about the data remain
    - Relationship between variables
    - Best practices for merging together and organizing CE files
  - Possible updates that can improve the usefulness of the data for future researchers
    - More subtlety in existing loan and expenditure variables