Welcome and Getting Started with Lifecycle Deficit Profiles

Gretchen Donehower

NTA Hands-On Workshop

Berkeley, CA

January 12, 2009

Plan

- Goals, Schedule, Expectations
- NTA Basics
- Getting Started
- Calculating Lifecycle Deficit Profiles

Goals

- Estimate a full set of profiles for each country
- Have all NTA researchers understand the project
- Learn by doing and teaching
  - Data and calculation problems and solutions
  - Innovations from standard methods

Schedule

- Two Lecture Sessions (optional)
  - Hands-on may start during session
- One Hands-On Workshop Session
  - Different topic areas in different locations
    - New to NTA (Conference Room)
    - Transfers (Tea Room)
    - Asset-based Reallocations (Basement)
    - Other (Attic)
    - OR GATHER IN REGIONAL GROUPS?
  - Exception on Wednesday
    - Generational Accounting (Conference Room)
    - New to NTA (Basement)

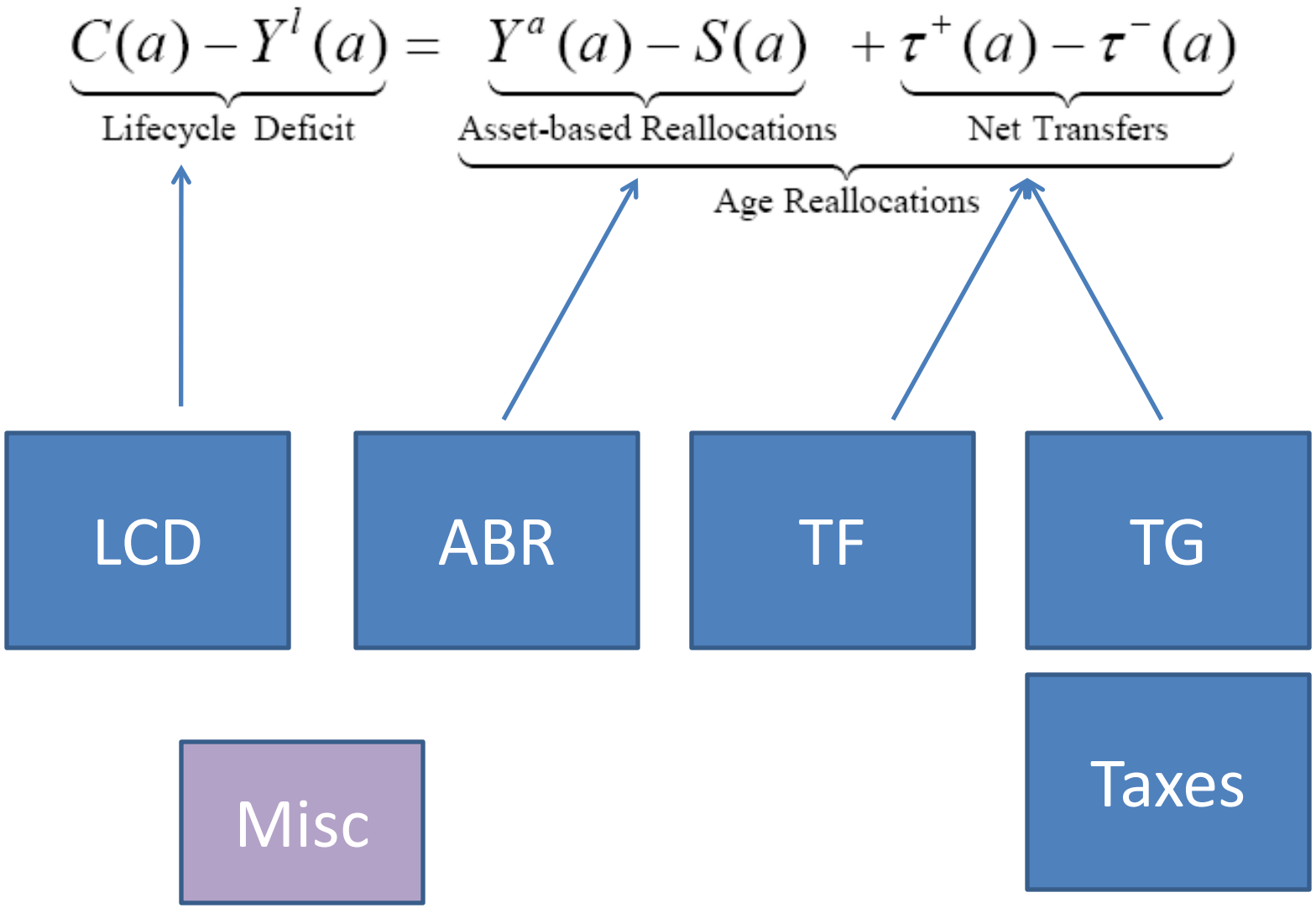
Schedule

- Friday
  - Morning session
    - Prepare “posters” of results and ongoing issues/questions
  - Midday session:
    - Circulate to observe other teams’ work
    - See Gretchen to upload latest estimates
  - Afternoon session:
    - Discuss issues, next steps
    - Celebrate!

Expectations

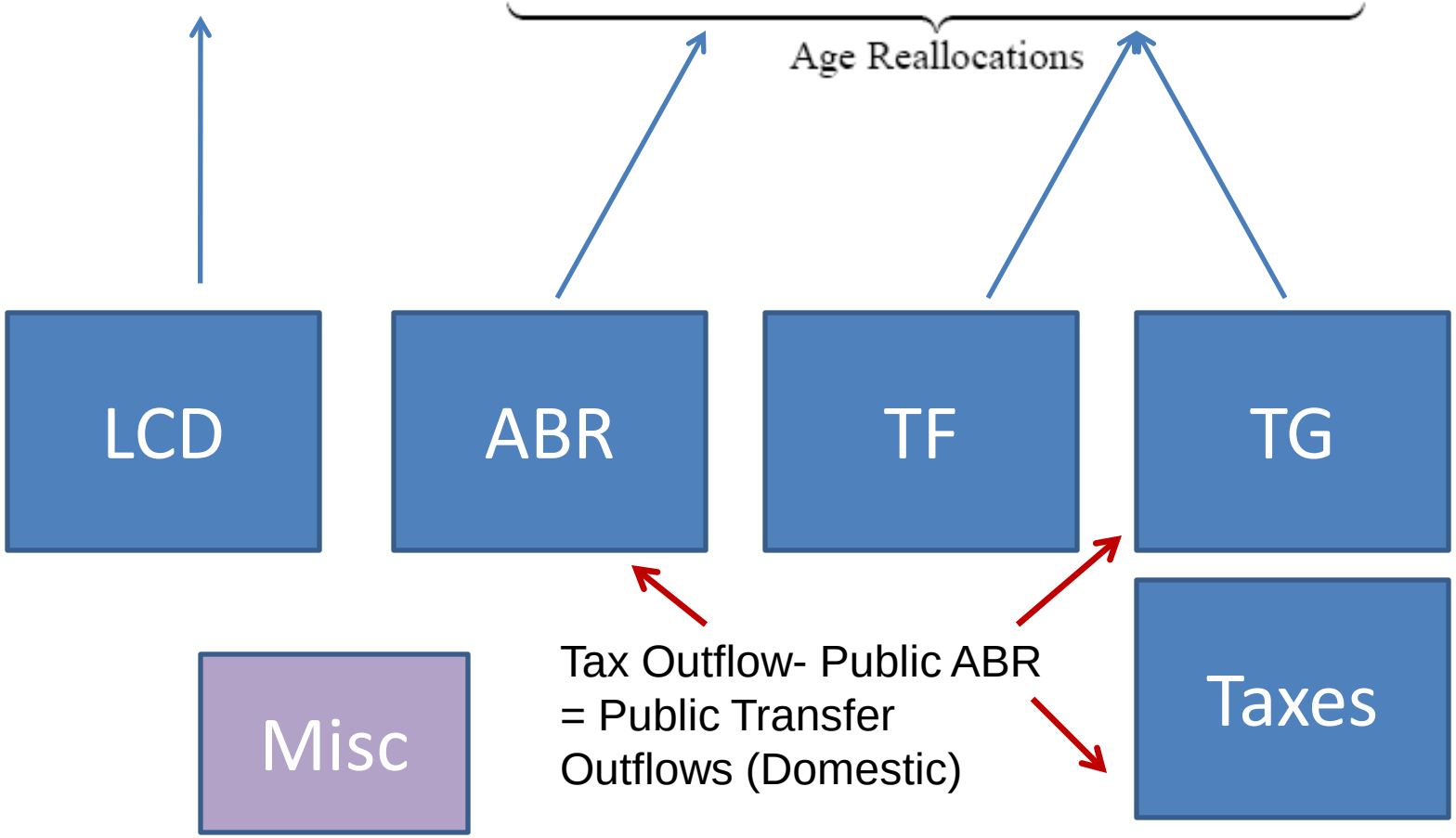
- Attend lecture sessions if you think they will be helpful, or work on your own
- Lots of interaction during lectures
- Help other teams
- NTA is a work in progress: nobody knows everything!

NTA Organization



NTA Organization

$$\underbrace{C(a) - Y^{l}(a)}_{\text{Lifecycle Deficit}} = \underbrace{\underbrace{Y^{a}(a) - S(a)}_{\text{Asset-based Reallocations}} + \underbrace{\tau^{+}(a) - \tau^{-}(a)}_{\text{Net Transfers}}}_{\text{Age Reallocations}}$$



NTA Basics – Two Steps

- Age shape estimated from surveys or administrative records
  - single years of age, per individual
- Control total adjusted
  - NTA aggregate matches the corresponding value from national accounts (with some modifications)

$$YLE_x^{\,adjusted} = YLE_x \left(\frac{YLE^{\,NationalAccounts}}{\sum_{x=0}^{\omega} YLE_x \times Pop_x} \right)$$

Other NTA Basics

- Nominal values
  - Eventually database will be able to give nominal or real values
- Annual
  - Based on annual national accounts
- Single units of local currency
  - For that time period, unless special case
- Smoothed profiles

NTA Basics – Age Shapes

- Administrative Data
  - Some public accounts
- Survey Data
  - Most profiles from survey data
  - Individual-level responses or household-level?
  - Entire population represented?
  - Sampling or other types of weights needed?
  - Survey quality issues

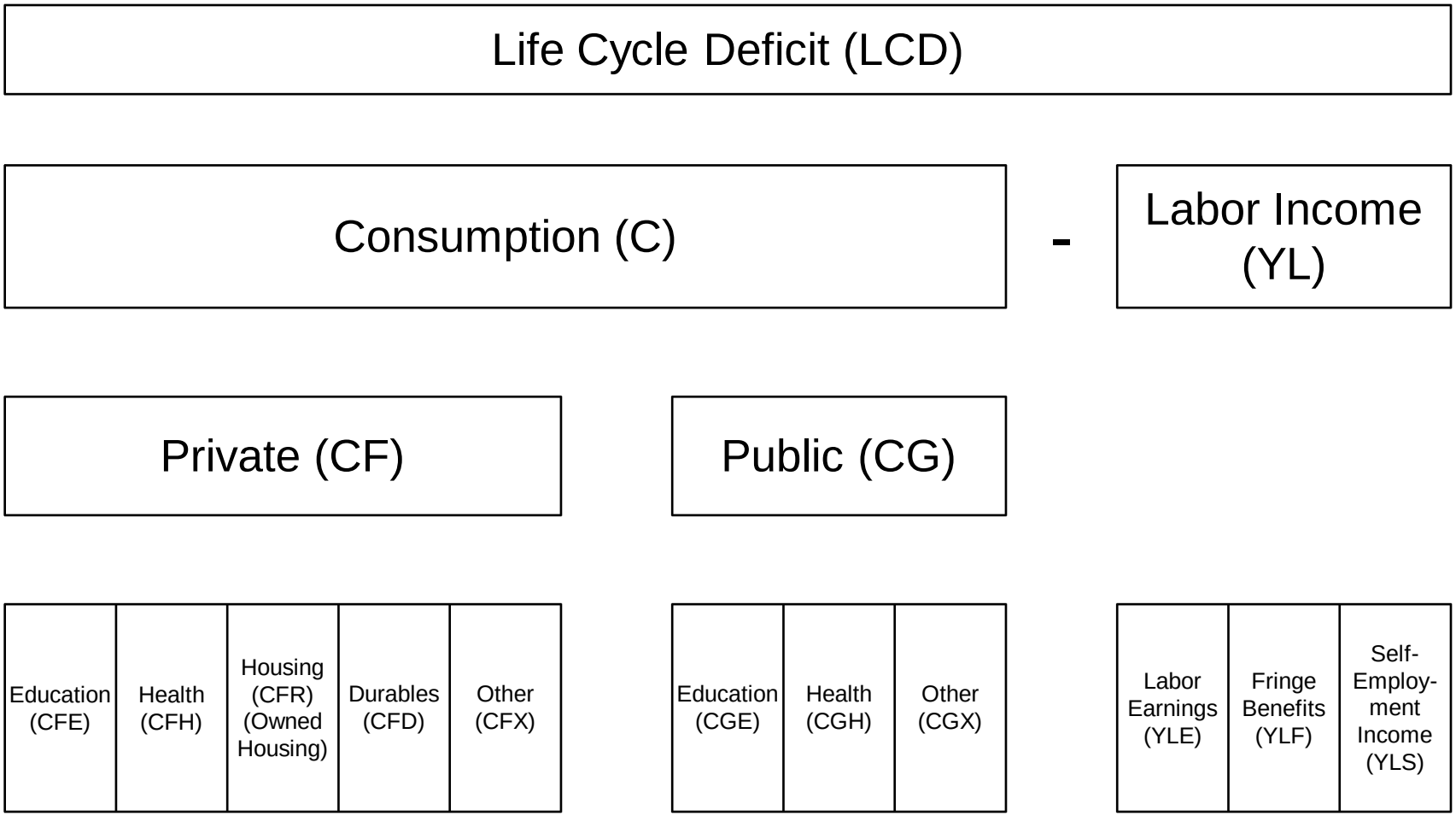
Survey Data Issues

- Missing data/inconsistent data/outliers
- Categorical vs continuous variables
- Top codes (especially of age)
- Nationally representative
- You must become an expert on the surveys you use!
  - How do other researchers deal with problems?
  - What other information can help solve the problems?

Looking Within the Household

- Some surveys have individual-level flows
  - Wages (YLE)
- Usually we must guess what goes on within the household
  - Equivalent Adult Consumer (EAC) Weights
  - Regression Methods
  - Iterative Methods

LCD Profiles



Private Education (CFE)

- Usually reported at household level
- Regression Method
  - Priv edu consumption of household *j* regressed on number enrolled (E) and non-enrolled (NE) of age *x*

$$CFE_j = \sum_{x=0}^{\omega} \alpha(x) E_j(x) + \sum_{x=0}^{\omega} \beta(x) NE_j(x)$$

  - Coefficients become within-household shares
- Alternative if you have no enrollment data
  - Consumption by level regressed on number in household by age, for restricted age ranges
  - Only appropriate in high-enrollment context?
- No smoothing!

Private Health (CFH)

- Usually reported at household level
- Method depends on data availability
  - If you have information on household members using in-patient (IN) or out-patient (OUT) services, method is similar to CFE

$$CFH_j = \sum_{x=0}^{\omega} \alpha(x) IN_j(x) + \sum_{x=0}^{\omega} \beta(x) OUT_j(x)$$

  - If you have external profile of per capita utilization by age (U) and number of household members by age (M)

$$CFH_j = \sum_{x=0}^{\omega} \beta(x) U(x) M_j(x)$$

    - Can go to higher-order specifications for a better fit if necessary

Private Health (CFH) cont.

- Iterative Method (non-parametric)
  - 1. Allocate household CFH equally to all household members
  - 2. Calculate the individual-level CFH profile
  - 3. Use the profile as shares to calculate a new allocation of CFH within each household
  - 4. Calculate a new individual-level CFH profile
  - Repeat steps 3 and 4 until the profile is stable

Survey Data:

HH	Per	Age	CFH	Iterations: 1		2		3		4		5		6		7	
				HH %	HH Amt	HH %	HH Amt	HH %	HH Amt	HH %	HH Amt	HH %	HH Amt	HH %	HH Amt	HH %	HH Amt
1	1	5	70	0.33	23.33	0.22	15.69	0.18	12.48	0.16	11.09	0.15	10.49	0.15	10.22	0.14	10.10
	2	35		0.33	23.33	0.39	27.16	0.41	28.76	0.42	29.45	0.43	29.76	0.43	29.89	0.43	29.95
	3	35		0.33	23.33	0.39	27.16	0.41	28.76	0.42	29.45	0.43	29.76	0.43	29.89	0.43	29.95
2	1	35	80	0.33	26.67	0.35	27.69	0.36	28.72	0.37	29.35	0.37	29.69	0.37	29.85	0.37	29.93
	2	35		0.33	26.67	0.35	27.69	0.36	28.72	0.37	29.35	0.37	29.69	0.37	29.85	0.37	29.93
	3	65		0.33	26.67	0.31	24.62	0.28	22.56	0.27	21.30	0.26	20.63	0.25	20.30	0.25	20.14
3	1	65	40	0.50	20.00	0.50	20.00	0.50	20.00	0.50	20.00	0.50	20.00	0.50	20.00	0.50	20.00
	2	65		0.50	20.00	0.50	20.00	0.50	20.00	0.50	20.00	0.50	20.00	0.50	20.00	0.50	20.00
4	1	5	20	0.50	10.00	0.50	10.00	0.50	10.00	0.50	10.00	0.50	10.00	0.50	10.00	0.50	10.00
	2	5		0.50	10.00	0.50	10.00	0.50	10.00	0.50	10.00	0.50	10.00	0.50	10.00	0.50	10.00

Resulting Profile:

5	14.44	11.90	10.83	10.36	10.16	10.07	10.03
35	25.00	27.42	28.74	29.40	29.72	29.87	29.94
65	22.22	21.54	20.85	20.43	20.21	20.10	20.05

Private Health (CFH) cont

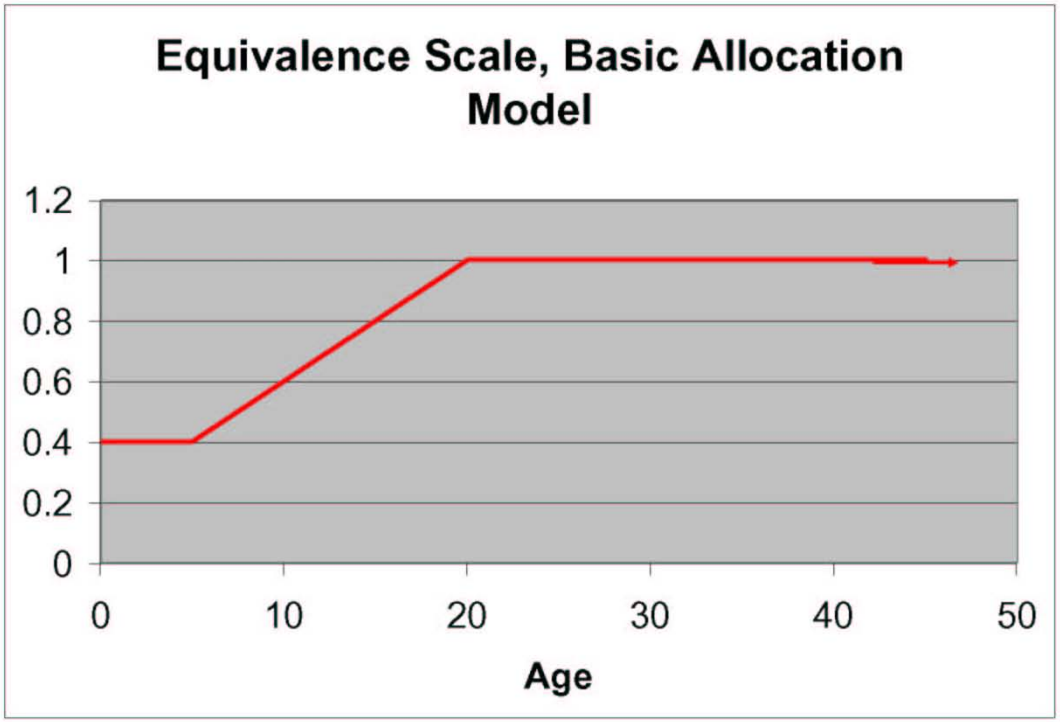
- Simple regression (old method)

$$CFH_j = \sum_{x=0}^{\omega} \beta(x) M_j(x)$$

  - This is now the least recommended approach
    - It can generate negative coefficients
    - Iterative method seems to give better estimates without requiring more data
  - Replace negative coefficients with zeros

Private Housing (CFR)

- Refers to owned housing only; rental housing is part of other consumption (CFX)
- Surveys often have “rental equivalent of owned home” or “imputed rent”
- Want to measure the flow of services from an owned asset, not the purchase price
- Household-level imputed rent is allocated to individuals using Equivalent Adult Consumer (EAC) weights



gen perweight = 1 – 0.6* (age>4 & age<20) * ((20-age)/16) – 0.6 * (age<=4)
egen hhweight = sum(perweight), by (hhid)

Then… CFR for each individual is (perweight/hhweight) * Household CFH

Private Durables (CFD)

- Household amount allocated with EAC weights
- Would like a similar flow value, as for owned housing
- Data is not available, though
- US uses purchase price and hopes that symmetric errors result in a valid age profile
- Any other bright ideas?

Private Other (CFX)

- This should be everything else from consumption categories
- Allocated using EAC Weights

Public Health and Education (CGH, CGE)

- Administrative records of usage by age are often available
  - Enrollment rates and average cost-per-student
- Specialized Surveys
  - Health surveys
  - Surveys for institutionalized population
- Make sure to include long-term care consumption
- Will depend on how much government measures

Public Other (CGX)

- Should be non-age-allocated expenses, so divide equal per-capita shares
  - Includes administrative expenses of age-allocated programs
- Some countries have data on age-allocated expenditures on other than health and education?

Labor Earnings and Fringe Benefits (YLE, YLF)

- YLE is wages and salaries
- Often available in surveys for individuals instead of households
- Usually good quality
- YLF is employer-paid health insurance, pension contributions and social protection insurance
- If data on YLF is not available, can assume it is a constant proportion of YLE

Self-Employment Income (YLS)

- Labor's estimated share of mixed income (income of unincorporated businesses)
- Amounts usually estimated at household-level, or individual-level estimates are suspect
  - Allocate based on relative shares of YLE profile
- Are there new methods being used for this? Like the iterative method?

Whew! Now that we have age shapes…

- Use weights to calculate unsmoothed means
- Smooth
- Adjust to control totals
- Make higher level variables
- Fill out LCD spreadsheet with documentation!

Smoothing

- Want to eliminate noise while preserving real features of the data
  - Can't smooth education because consumption is too tightly age defined
  - Need relative age group weights
- Use SUPSMU in R
  - For many profiles, standard "CV" span is fine
  - For others, like YLE and anything with sharp "elbows", need to experiment with short smoothing spans, or dropping some ages from the smoother

Making R and Stata Play Nicely

- Within Stata program:

```
use unsmoothed_yle_profile, clear
gen skip=1 if yle==0 and age<20
gen varnm="yle"
outsheet age yle weightvar  using ~/ntadata.csv if
        skip~=1, comma nolabel replace
shell R --vanilla < supsmu.r > supsmu.out
insheet using ~/smoothedvar.csv, comma nonames clear
rename x age
rename y yle_sm
keep age yle_sm
```

Making R and Stata Play Nicely

- R program:

```
ntadata<-read.csv("~/ntadata.csv",header=FALSE)
whichspan<-ntadata[1,4]
spanval<-NULL
if (whichspan=="yle") spanval<-0.04
if (whichspan=="cfx") spanval<-"cv"
test<-supsmu(x=ntadata[,1],y=log(ntadata[,2],
        wt=ntadata[,3],span=spanval)
test$y<-exp(test$y)
write.csv(test,"~/smvar.csv")
```