Ethiopia Poverty Measurement Training

Day 5: deflators

Deflators

- Need to adjust nominal consumption to reflect the actual prices paid by households
- Prices can vary over time due to
  - general inflation
  - seasonal variation
  - only a concern when data collection takes place over a longer period (>2 months?)
- Prices can across space
  - often higher in urban areas

Various Price Indices

	Paasche	Laspeyres	Fisher
Textbook Formula	$P^h = \frac{p^h \cdot q^h}{p^0 \cdot q^h}$	$L^h = \frac{p^h \cdot q^0}{p^0 \cdot q^0}$	$F^h = \sqrt{P^h L^h}$
D&Z Formula	$P^h = \left(\sum w_k^h \frac{p_k^0}{p_k^h} \right)^{-1}$	$L^h = \sum \left(w_k^0 \frac{p_k^h}{p_k^0} \right)$	

- $\boldsymbol{w_k^h}$ is the share of good k in the household's total consumption: $p_k^h \cdot q_k^h / \sum_k p_k^h \cdot q_k^h$
- $\boldsymbol{w_k^0}$ is the share in total national consumption: $p_k^0 \cdot q_k^0 / \sum_k p_k^0 \cdot q_k^0$

Price Sources

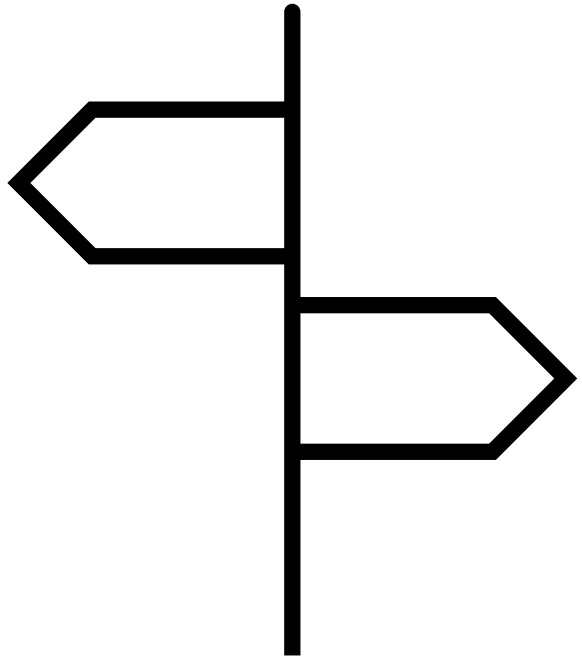
Source	Pros	Cons
Unit values (from observed hh purchases)	• Collected in exactly the areas as the households in the survey • Classification of foods aligns perfectly	• Better-off households may consume higher quality versions of items • Within each classification, households will tend to consume more of what is locally relatively cheaper • Usually only available for food
Price survey	• Collected in exactly the areas of the households in the survey • Classification of items can align perfectly • Precise quality definitions can be applied • Covers food and nonfood	• Rarely implemented, additional cost • Needs to be done carefully, otherwise will just introduce noise
Raw CPI data	• Data on large number of items • Should have consistent quality definitions • Available at no extra cost • Covers food and nonfood	• Often only collected in urban areas • May not have access to raw data • Coding of items may be different from survey

Other Decisions

- Use separate temporal and spatial price deflators, or create one joint price index?
  - I recommend using a joint one UNLESS you want to use the CPI for temporal deflation
- If using separate ones, which one to apply first?
  - M&V recommend applying temporal first
- Use separate deflators for separate components of the consumption aggregate (food vs. non-food, by top-level COICOP)?
  - M&V recommend against this: can distort share of food in total consumption
- Use lower level of aggregation for budget shares than for prices?
  - Budget shares can be down to household level. Prices can be down to cluster-level.

Three Good, Common Options for Deflators

1. Joint spatial-temporal Paasche index. Add nonfood prices if you have good data from a source such as market survey or raw CPI data. (Nigeria)
2. Deflate temporally using official monthly CPI (at regional level if available). [Do this adjustment before aggregating unit values to value consumption.]
   a. Deflate spatially using spatial Paasche index. Add nonfood prices if possible.
   b. Deflate spatially using implicit spatial deflators based on relative poverty lines. (EHCVM1 countries)



Spatial / Temporal Concerns

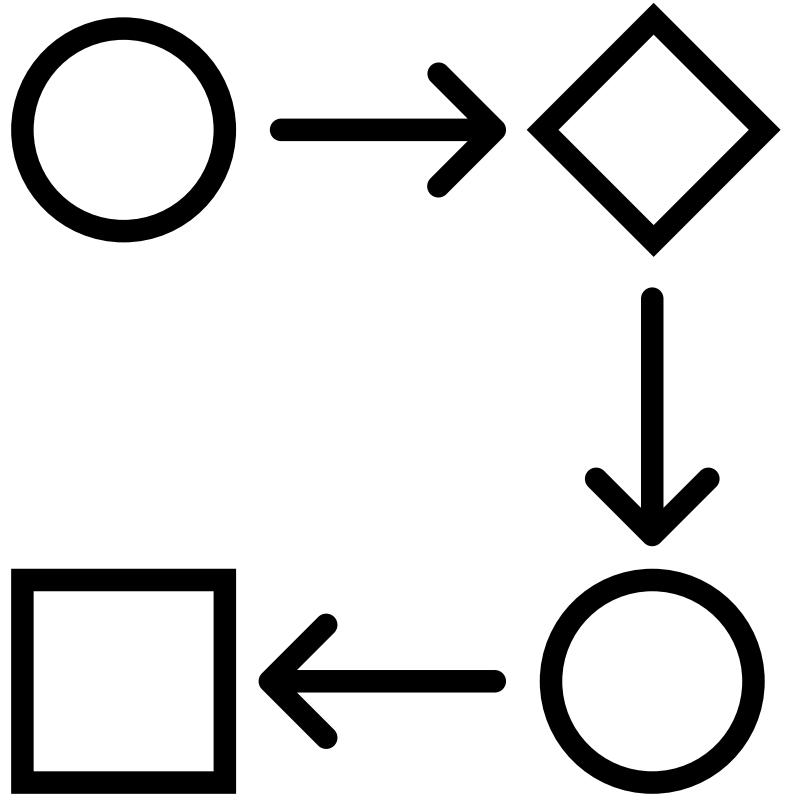
- A price index at the PSU level is by definition a joint spatial-temporal price index, as each PSU is only interviewed at one point in time
  - thus do NOT also apply a temporal deflator based on the CPI or anything else
- If you want to use the CPI for temporal deflation, the spatial deflator must be at a domain such that the sample is balanced across the domain each month
  - or do temporal deflation with a quarterly index if your sample is only balanced by quarters

```
. tab my region, nofreq row
```

my	IHS5 2019 Region location			
	North	Central	Southern	Total
2019-04	37.14	25.71	37.14	100.00
2019-05	16.76	35.32	47.93	100.00
2019-06	18.73	34.66	46.61	100.00
2019-07	17.89	33.50	48.60	100.00
2019-08	19.45	35.13	45.42	100.00
2019-09	19.09	37.52	43.39	100.00
2019-10	18.63	31.78	49.59	100.00
2019-11	7.54	41.63	50.83	100.00
2019-12	19.06	32.21	48.73	100.00
2020-01	19.46	34.85	45.69	100.00
2020-02	25.27	29.89	44.84	100.00
2020-03	22.09	38.56	39.35	100.00
2020-04	20.16	20.69	59.15	100.00
Total	19.03	34.56	46.41	100.00

Deflators – Paasche or Laspeyres from D&Z formulas

1. [Strategy for different item-unit prices]
2. Construct base price dataset (p^0)
3. Construct weights = budget shares (for each item or item-unit)
4. Aggregate budget share * relative prices
5. Apply to nominal consumption to get real consumption



Step 1: Strategy for Different Item-Unit Prices

- If we have separate prices / unit values for each unit for each item:
  - Can leave as is and construct weights for each item-unit combination
  - Could use only prices for most common unit for each item
  - Can combine in Jevon’s index (geometric mean of relative prices)

Step 2: Construct Base Price Dataset

- We have p^h which are local[-current] prices at some level, say region
- Take the sum of the sampling weights over each region
- Use this to construct p^0 as the weighted mean of p^h
- This is better than just taking the overall national median price
  - Tend to have more price observations in urban areas, so overall national median price is biased to urban areas
  - Using weighted mean will make a Laspeyres price index average to exactly 1
- Constructing p^0 as national price (rather than using prices in specific domain-period) minimizes the differences between p^h and p^0, making it a better approximation for the true-cost-of-living

Step 3: Construct Weights = Budget Shares

- The Paasche price index formula $P^h = \left(\sum w_k^h \frac{p_k^0}{p_k^h}\right)^{-1}$ uses household-specific budget shares and thus gives a different value of the index for every single household
  - This is what is consistent with adjusting a money metric utility function
  - We should have pretty accurate information on household budget share
  - But: definitely check the distribution of this index and winsorize outliers
- In practice, the weights are often constructed at a higher level
  - Often just use weights at the same level as p^0
- For Laspeyres index, just take national budget share of each item [item-unit]

Step 3: Construct Weights = Budget Shares

- The Paasche price index formula $P^h = \left(\sum w_k^h \frac{p_k^0}{p_k^h}\right)^{-1}$ uses household-specific budget shares and thus gives a different value of the index for every single household
  - This is what is consistent with adjusting a money metric utility function
  - We should have pretty accurate information on household budget share
  - But: definitely check the distribution of this index and winsorize outliers
- In practice, the weights are often constructed at a higher level
  - Often just use weights at the same level as p^0
- For Laspeyres index, just take national budget share of each item [item-unit]

monthname	headline_cpi	temporal deflator
2018m1	101.30	0.994744
2018m2	102.09	1.002459
2018m3	101.18	0.993525
2018m4	102.19	1.003449
2018m5	103.25	1.01388
2018m6	104.47	1.025879
2018m7	103.95	1.020791
2018m8	100.50	0.986894
2018m9	100.30	0.98493
2018m10	100.90	0.990822
2018m11	100.70	0.988858
2018m12	101.20	0.993768
	101.83	

CPI-based Temporal Deflators

- Calculate average CPI level over all the months of fieldwork (using the share of fieldwork per month if it varies)
- For each month, divide the monthly CPI level by the average CPI level.
- Use this as temporal deflator

Implicit Spatial Deflators - Method

- Construct basket of food items as for poverty line
- Value this basket at domain prices
- Gives a Laspeyres spatial price index for food

OR

- Construct basket of food items as for poverty line separately for each domain
- Value each domain basket at domain prices
- Gives a Paasche spatial price index for food

THEN

Add a nonfood component for each domain using the Ravallion method (covered Friday)

Gives domain-level poverty line

Divide domain-level poverty line by national poverty line for implicit spatial deflator

Implicit Spatial Deflators - Discussion

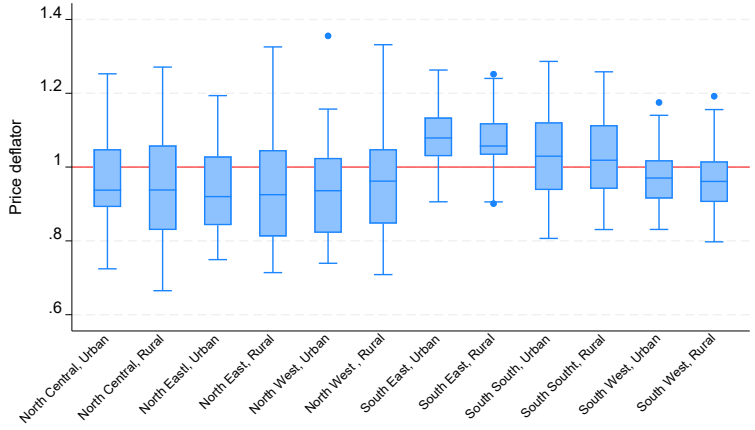
- Starts with a food price index and adds a nonfood component without needing nonfood prices
- Domain level basket as better if food consumption patterns vary significantly across the country
  - particularly if some key foods are staples in one area, but rarely consumed in another
- Combines nicely with using official CPI for temporal deflation

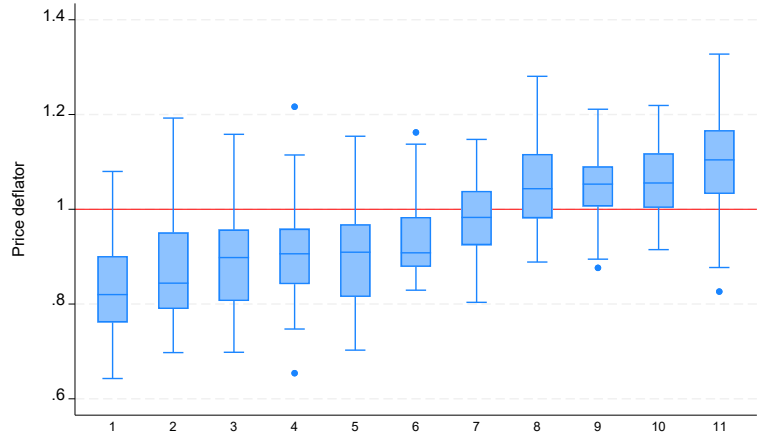
Robust Construction of Deflators

- Check price ratios are within a reasonable range (say 0.5 to 2 if prices in one part of the country are not less than half or more than twice the national price average)
- If constructing household or PSU-level indices, Winsorize top and bottom outliers
- If constructing higher level price indices, inspect and make sure they are reasonable
- Check that the weighted average of deflators is close to 1 (if using weighted national price as p^0)

Nigeria

- Very disaggregated deflator (PSU level?)
- Overall range is quite reasonable
- Clear
- increase in average over time
- Some variation between different parts of the country
- Little urban-rural difference





Aggregation and Final Checks

Household-Item level data and the final NCA

Household-Item Level Data

- We have consumption of each item, food and nonfood, including use value of durable goods and housing rental value
- Very helpful to maintain this as a final dataset that can be used for different types of analysis in the future
- Can add COICOP coding (1 to 3 digits) to this dataset
- Can also include source of consumption: purchase, own production, gift, use value etc.
- In general, these are nominal, annual values

Aggregating and Final Checks

- Sum up consumption of all items over the household
- Drop households for which the NCA cannot be constructed
  - households missing entire sections of the questionnaire
  - households deemed to have too many missing values / imputations otherwise
  - household with extreme share of food in total consumption
- Reweight to adjust for dropped households
  - rescale at PSU level so total weight for PSU remains the same
- Apply spatial and / or temporal deflators
- Adjust for household size and composition
- Look at overall distribution
  - the Stata command `outdetect` is useful (doesn't assume log-normal distribution)
  - look at food and nonfood components separately
  - both total and per capita household consumption
  - try to understand and address cause of outliers (rather than just dropping or winsorizing)