Training session on Income and Wealth Measurement in Household Surveys

Session 3. Adjusting for prices and household size composition

Nov 10, 2022

Dean Jolliffe,

Samuel Tetteh-Baah, Sergio Olivieri

Slides prepared for the training session as part of the joint IARIW-TNBS Conference on Measurement of Income, Wealth and Well being in Africa (Nov 11-13, 2022). Several slides are drawn from the work of Sergio Olivieri and Samuel Tetteh-Baah.

Financial support from the UK Government through the Data and Evidence for Tackling Extreme Poverty (DEEP) Research Programme is gratefully acknowledged.

THE WORLD BANK
IBRD • IDA | WORLD BANK GROUP

Presentation outline

I. Spatial price adjustments for income and wealth
   i. Brief overview
   ii. US example
   iii. Comments

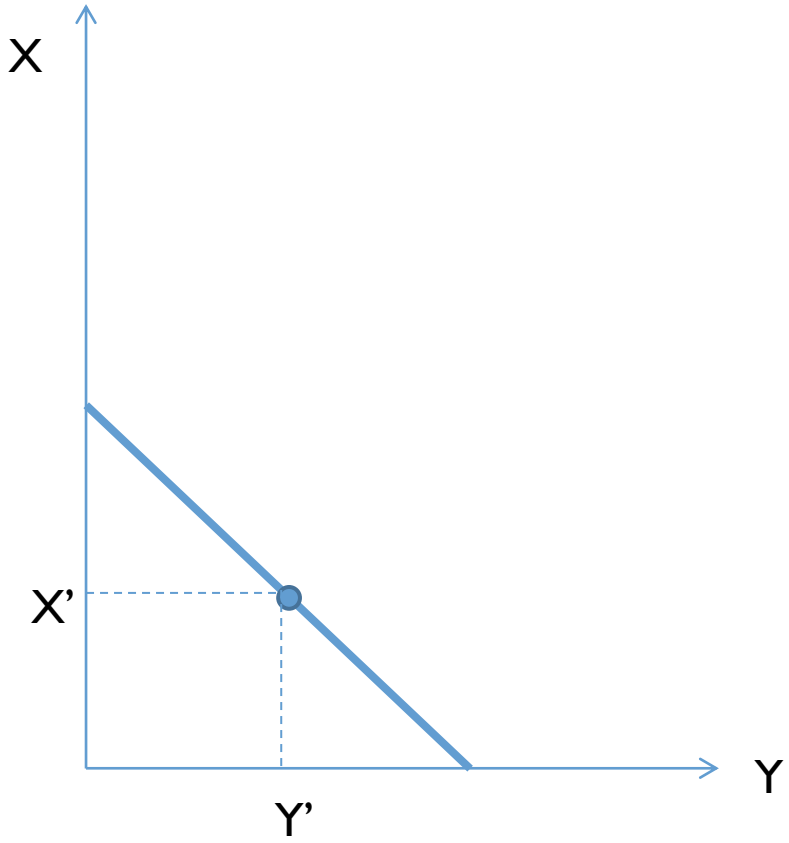
II. Adjusting for household composition -- Adult-equivalence and household economies of scale
   i. Brief overview
   ii. Global example – household economies of scale
   iii. Comments

III. Tangent on asset indices
   i. Overview
   ii. Shortcomings

Adjustments: Spatial Price Differences

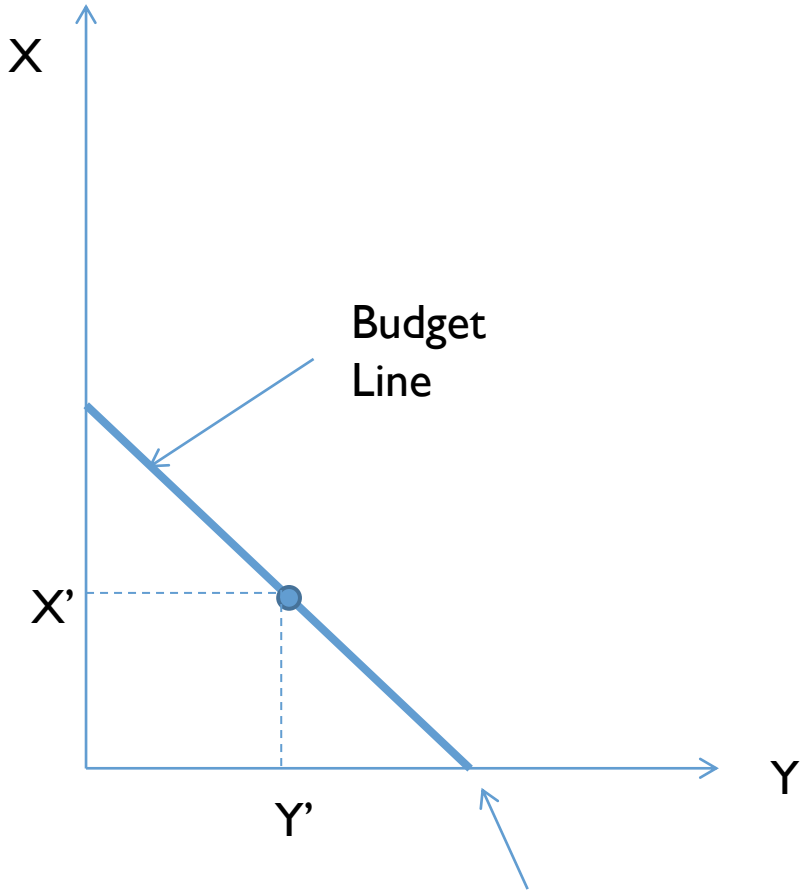
Concepts, Example, Comments

Let's go back to introductory economics



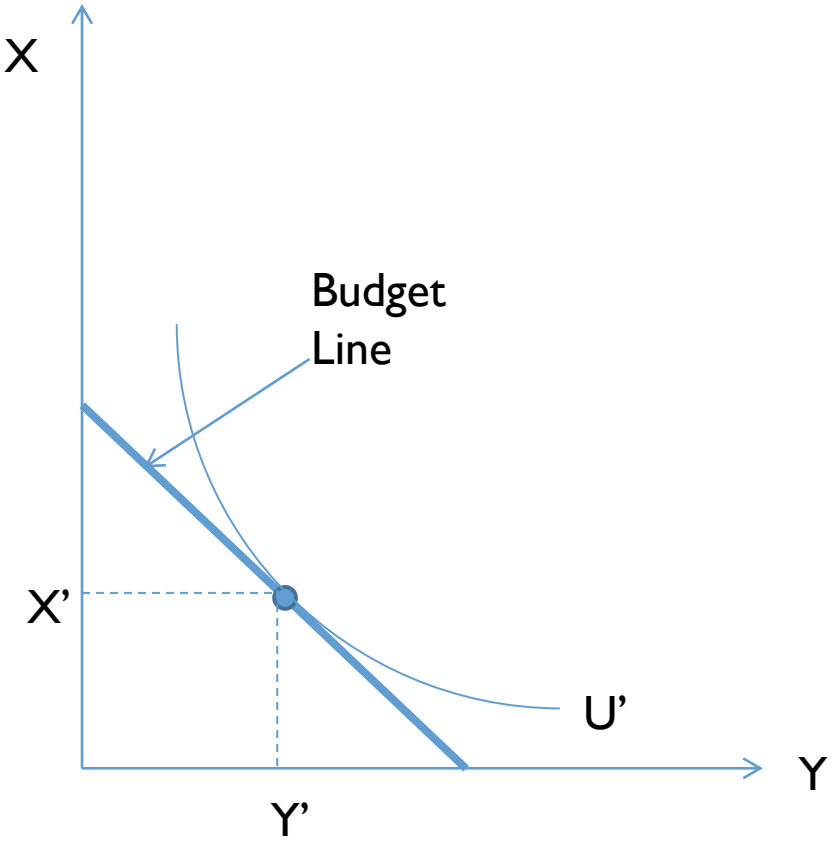
WORLD BANK GROUP

Let's go back to introductory economics

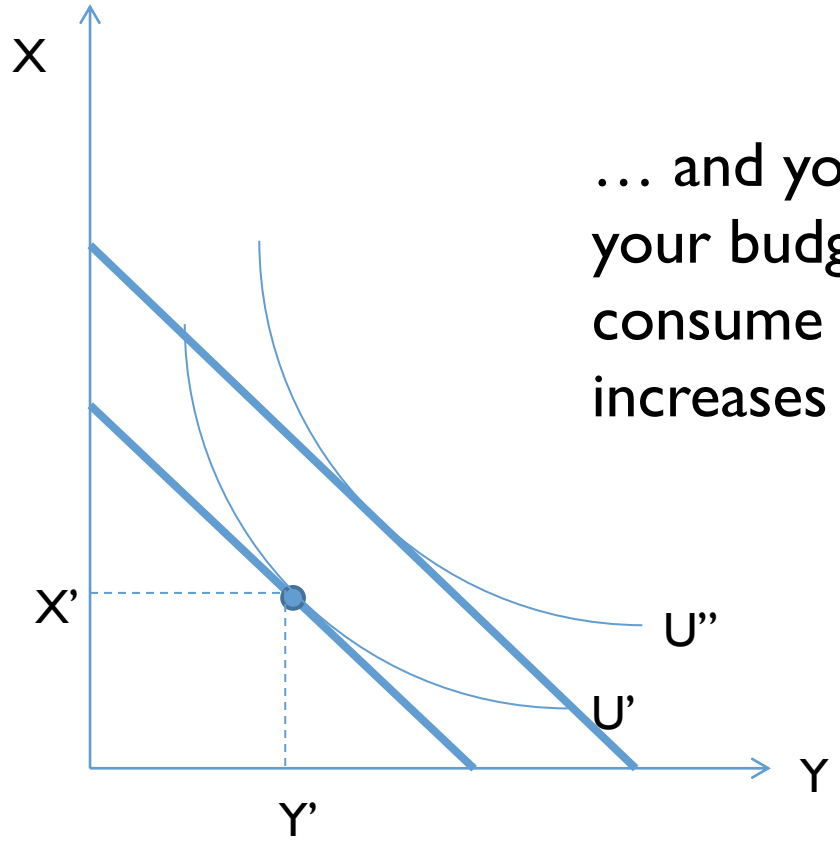


If you only want y, you can consume …

Let's go back to introductory economics

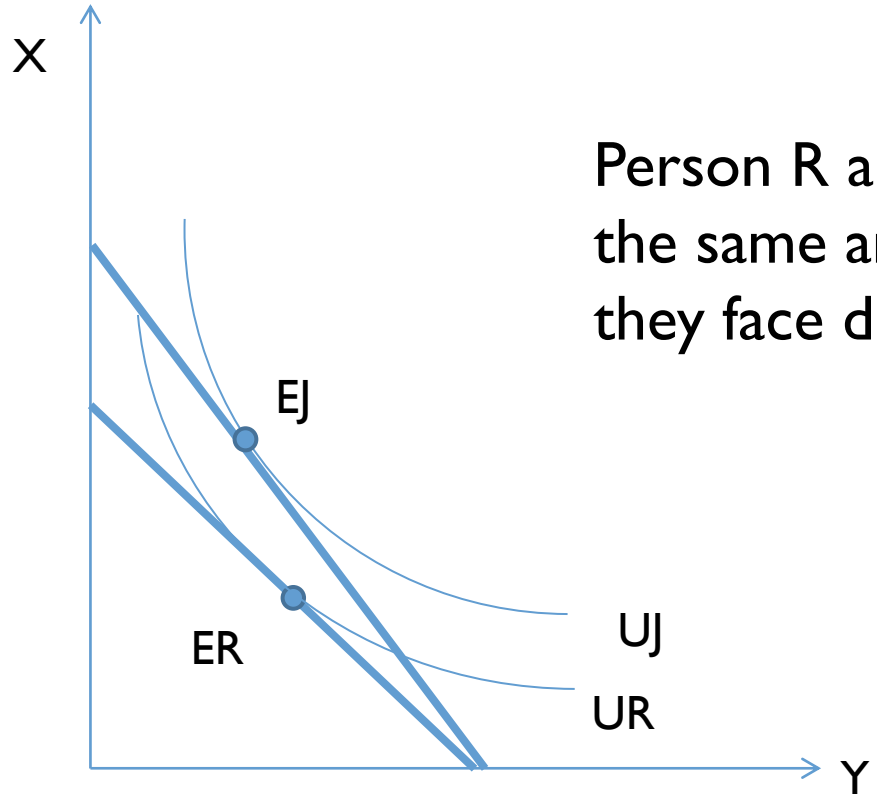


If prices are the same …



… and your income increases, your budget expands and you can consume more and your utility increases

If prices differ …



Person R and Person J spent exactly the same amount (say, $100) but they face different prices

ER=EJ(=$100)
but UR<UJ

Who faces lower prices?
Who is better off?

A simple example: who is better off?

	Consumption (q)		Price ($/kg)		Consumption Nominal ($)
	Rice (kg)	Meat (kg)	Rice	Meat	
Ringo	2	2.5	10	32	100
John	4	2.8	4	30	100
Paul	4.4	3.0	5	26	100
George	15	7.0	2	10	100

Who is better off?

Who is better off?

Ringo:_pR: 2*10 + 2.5*32 = 100
Ringo_pJ: 2*4 + 2.5*30 = 83 + more

WORLD BANK GROUP

Paasche vs Laspeyres (in spatial price adjustment)

- In the context of spatial price adjustments, **Paasche** price index means comparing two price regimes using individual *i-th* consumption

$$P_P = \frac{P_i Q_i}{P_R Q_i}$$

- **Laspeyres** price index means comparing two price regimes using a fixed consumption of reference

$$P_L = \frac{P_i Q_R}{P_R Q_R}$$

- Basically, weights for prices are different

Two ways to make Ringo's expenditure comparable to John's

Calculate Ringo's real expenditure					
weights	Evaluate it at John's prices (denominator)	Evaluate it at Ringo's prices (numerator)	Price index	Ringo's real expenditure	Notes
Ringo's consumption	83	100	1.20	83	Paasche
John's consumption	100	130	1.30	77	Laspeyres

To report Ringo's real expenditure (as opposed to nominal) we can account for the fact that Ringo faces higher prices by comparing his nominal exp with how much he would have had to pay if he faced John's prices. Alternatively, we could compare Ringo's price relative to John's price using John's consumption level as the reference point. (if Ringo were purchasing John's bundle, how much more would he need)

Note: 100/83 = 1.2 and 130/100 = 1.3
The selection of index affects Ringo's real expenditure!

Comparison of all four players (Reference prices=John's)

- Paasche Price Index

	Each person's quantity at John's prices	Each person's quantity at own prices	P_P	Real E
Ringo	83.0	100	1.2	83.0
John	100.0	100	1.0	100.0
Paul	107.6	100	0.9	107.6
George	270.0	100	0.4	270.0

$$P_P = \frac{P_i Q_i}{P_J Q_i}$$

- Laspeyres Price Index

	John's quantity at John's prices	John's quantity at each individual own prices	P_L	Real E
Ringo	100.0	129.6	1.3	77.2
John	100.0	100.0	1.0	100.0
Paul	100.0	92.8	0.9	107.8
George	100.0	36.0	0.4	277.8

$$P_L = \frac{P_i Q_J}{P_J Q_J}$$

What's your best estimate of total household expenditure?
Spatial price adjustments important to correctly sort people on wellbeing, but be careful if these data are being used for cross-country comparisons.

WORLD BANK GROUP

Comparison of all four players (Reference prices=John's)

- Paasche Price Index

	Real E (Paasche)	Ratio of sums	Scaler	Real E scaled to national average
Ringo	83.0	560.6/400	1.4015	59
John	100.0	560.6/400	1.4015	71
Paul	107.6	560.6/400	1.4015	77
George	270.0	560.6/400	1.4015	193

$nominal\ sum: 400$
$real\ sum: 400$

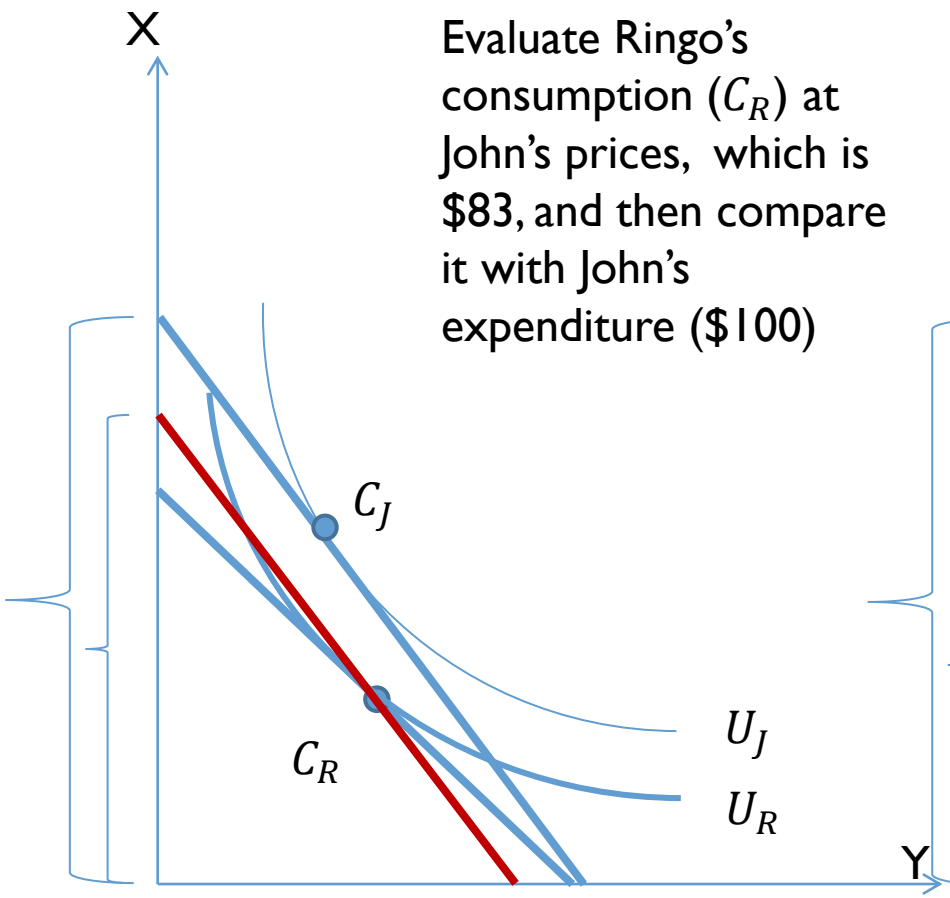
Note that scaling did not change the dispersion, relative positions unchanged.

e.g., round(270/83)=3.3
round(193/59)=3.3

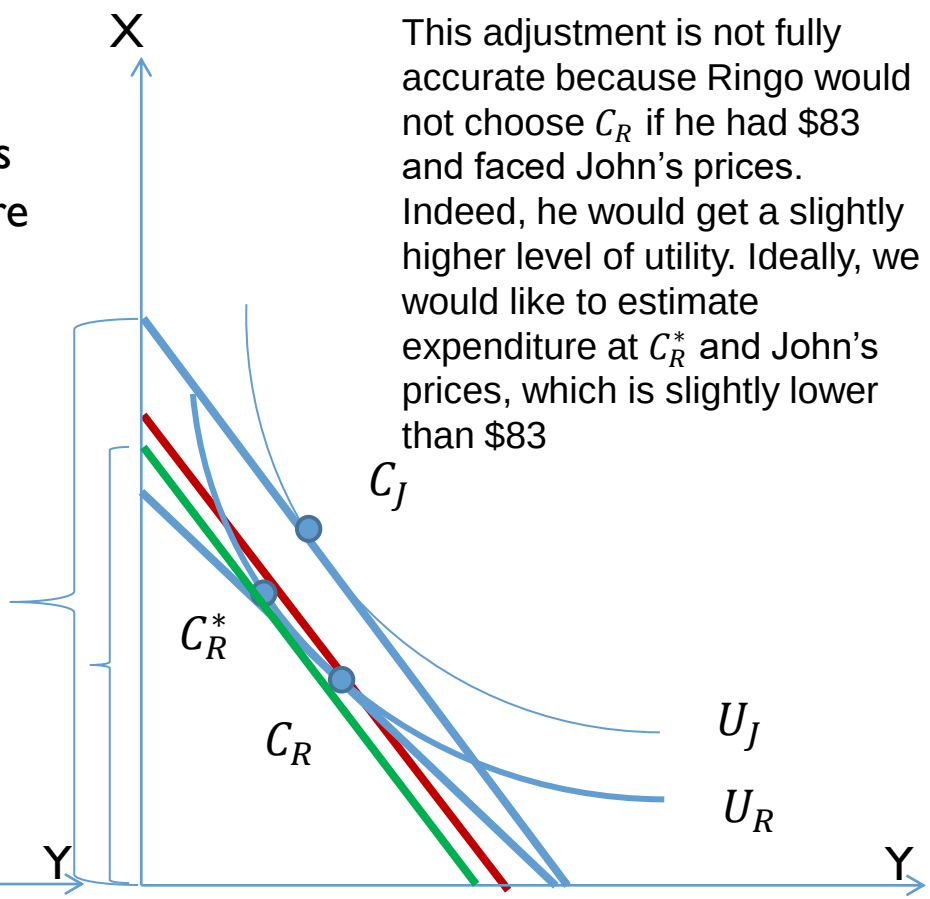
WORLD BANK GROUP

What's your best estimate of total household expenditure?
Spatial price adjustments important to correctly sort people on wellbeing, but be careful if these data are being used for cross-country comparisons.

Visual presentation of Paasche Index



Evaluate Ringo's consumption (C_R) at John's prices, which is \$83, and then compare it with John's expenditure (\$100)



This adjustment is not fully accurate because Ringo would not choose C_R if he had \$83 and faced John's prices. Indeed, he would get a slightly higher level of utility. Ideally, we would like to estimate expenditure at C_R^* and John's prices, which is slightly lower than \$83

Ideal cost-of-living adjustment

- *The ideal cost-of-living index*

$$P^* = \frac{E(P_i, U_i)}{E(P_R, U_i)}$$

$$= \frac{How\ much\ you\ paid}{How\ much\ you\ would\ pay\ under\ reference\ prices\ to\ keep\ your\ utility}$$

- Paasche is an approximation of it.

- Laspayres is an approximation for

$$P^* = \frac{E(P_i, U_R)}{E(P_R, U_R)}$$

$$= \frac{How\ much\ reference\ person\ would\ pay\ under\ your\ prices}{How\ much\ reference\ person\ paid}$$

- Difference is whose utility will be the basis for the price adjustment

Comments:
Unit value vs. Prices, CPI vs. price survey

- Price data are collected from markets.
  - Usually identification of quality is quite rigorous
  - Often price data are collected only from select urban areas
  - Relatively reliable
- Unit value is a ratio of expenditure to quantity
  - Different qualities might be included in the same item
  - Unit values can be representative at sub-national level
  - Due to misreporting of units, it is often very noisy
- CPI vs. price survey, discuss in context of spatial price adj
  - Frequency
  - Coverage
  - Comparability

Unit values and quality

(i) If we can distinguish coarse and fine rice in quantity and prices

	Consumption (kg)		Price ($/kg)			
	Coarse Rice	Fine Rice	Coarse Rice	Fine Rice		
John	4	2.8	4	30		
George	15	7	2	10		

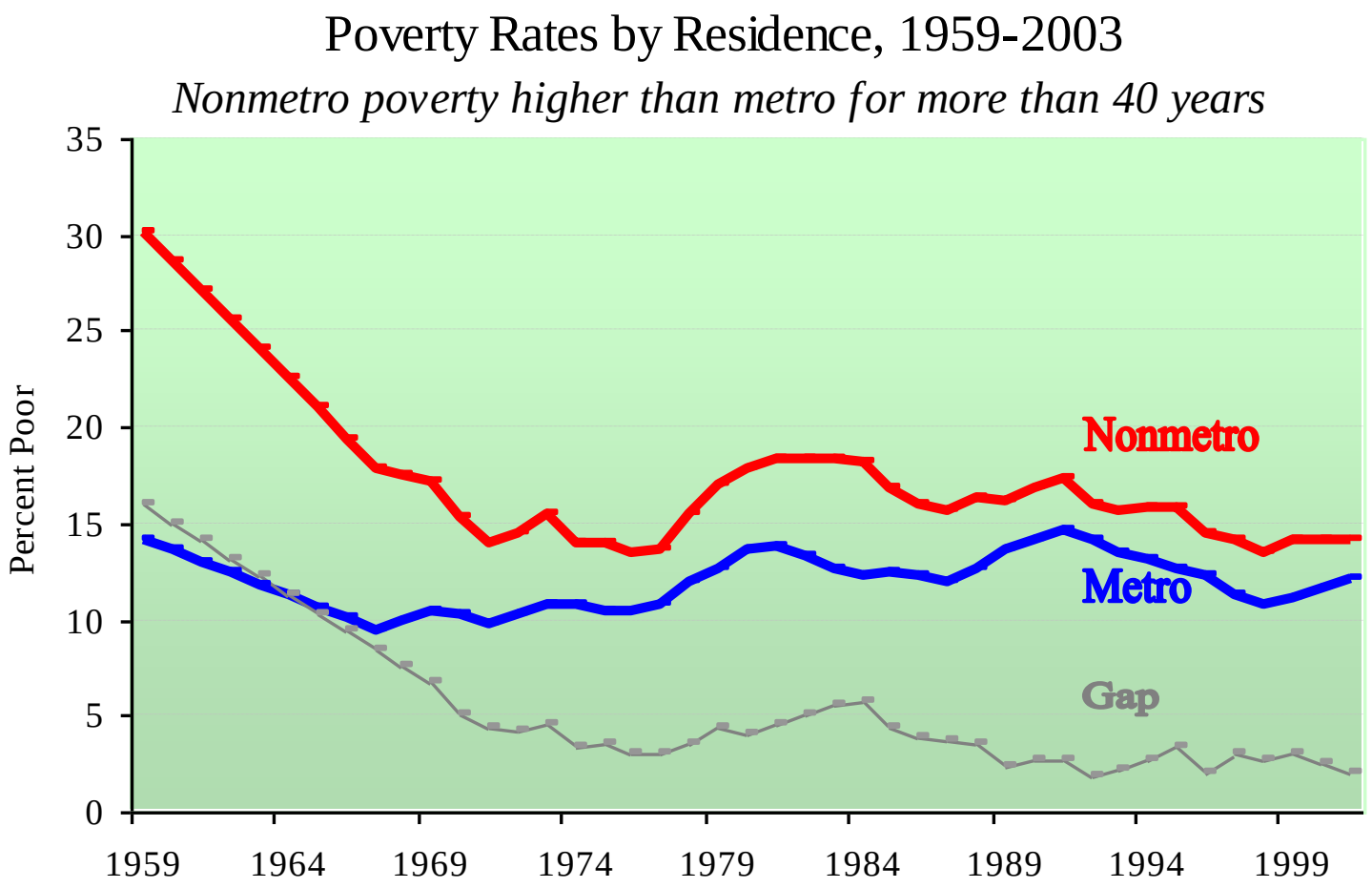
(ii) If we cannot distinguish coarse and fine rice and can observe only observe quantity and expenditure

	Consumption (kg)	Price ($/kg)	Unit Value ($/kg)
	Rice, qty	Rice, exp	Rice
John	6.8	100	14.7
George	22	100	4.5
Price Ratio			

Comments, tangential

- In the context of intertemporal price adjustments
  - Paasche index uses budget shares of the current period
  - Laspeyres index uses budget shares of the reference year
  - Fisher Ideal is often used; $P_F = \sqrt{P_P * P_L}$
- Discuss difference between spatial price index and a cost-of-living index
  - Ratio of poverty lines
- Compare national mean consumption adjusted and unadjusted
  - What if mean consumption adj > mean consumption nominal
  - Desirable to recenter consumption on nominal mean after adjustment

Example: Poverty in the U.S.



***Note*: Imputed for years 1960-69, 1970 and 1984. Metro-Nonmetro definitions changed in fourth year of decade.**

***Source*: Current Population Survey and Census Bureau's Consumer Income Series P-60 reports.**

Poverty is higher in nonmetro areas, and Federal assistance is targeted

- Persistently poor nonmetro counties receive block grants amounting to more than $1,000 per person.
- Per capita distribution of Federal funds for income security programs higher in nonmetro areas (17% higher in early 2000s)
- Nonmetro, per capita Food Stamp benefits > metro, per cap benefits. (more than 30% in early 2000s)

Brief Description of Fair Market Rent (FMR) Data

- In US, there is no adjustment for spatial differences in prices. There is no official data on this. But,
- Fair market rental (FMR)value is collected by government for housing voucher program
- FMR of rent + utilities for 'standard quality' housing (evaluated at 40th percentile)
- Full coverage (354 metro, 2350 nonmetro), aggregated to 100 observations (met/nonmet by State)

FMR Index

- FMR Index: $0.44*FMR_i + 0.56*Other$

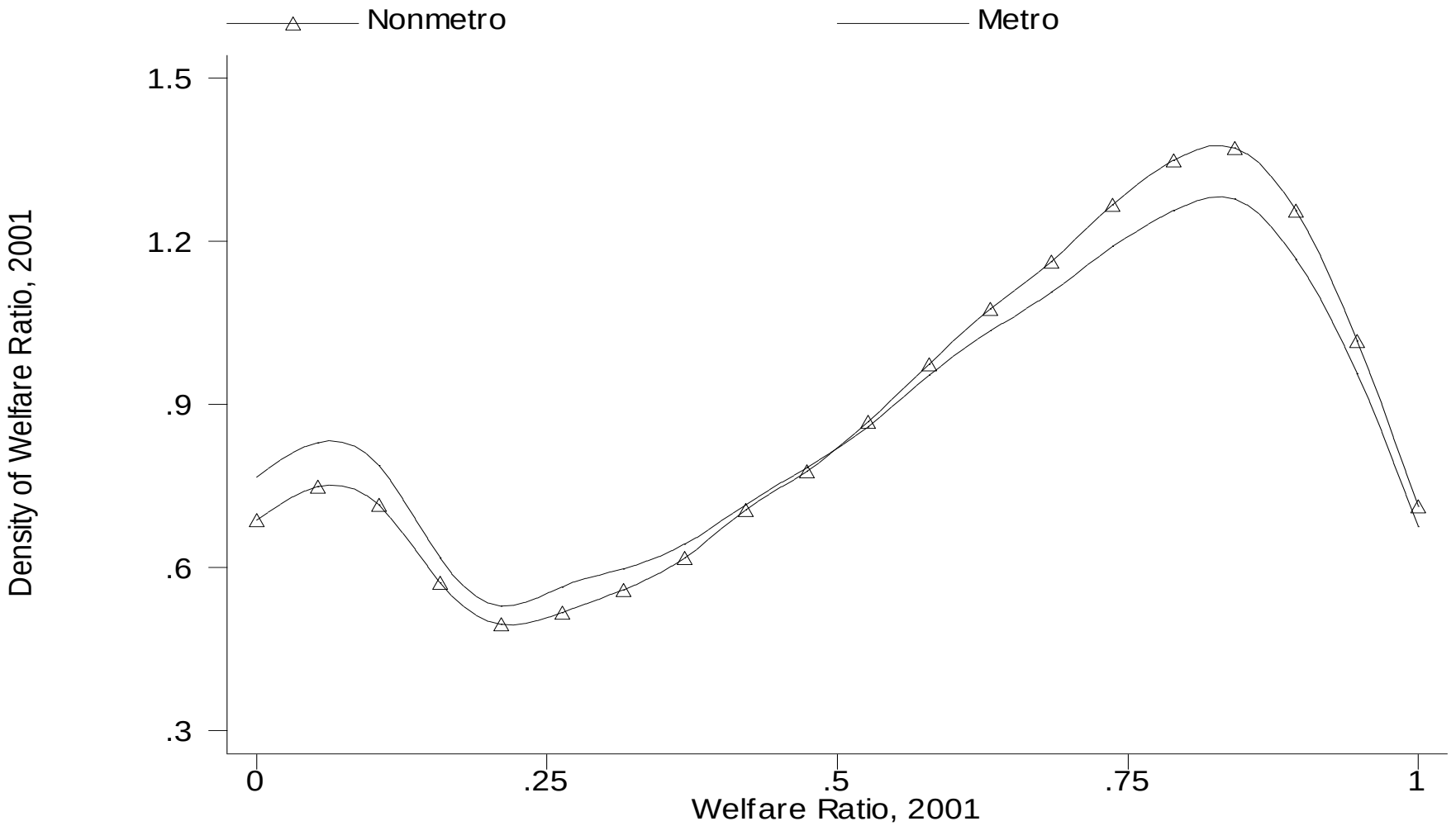
- Findings robust to weight change down to 33% (before difference is no longer stat. sig.)
- Findings robust to correlation between FMR and all other good prices ranging from (-0.2, 1).

Table 1: Scaled Fair Market Rent Index, Nonmetro-metro Comparison

Fair Market Rent Index	Average	Median	(Min, Max)
National	1.00	1.00	(0.74, 1.21)
Nonmetropolitan	0.82	0.81	(0.74, 1.21)
Metropolitan	1.04	1.01	(0.85, 1.19)

Notes: **Fair Market Rent index weighted by individual weights to match weights used for poverty estimation.**

Distribution of Income to Needs (Poor, 2001)

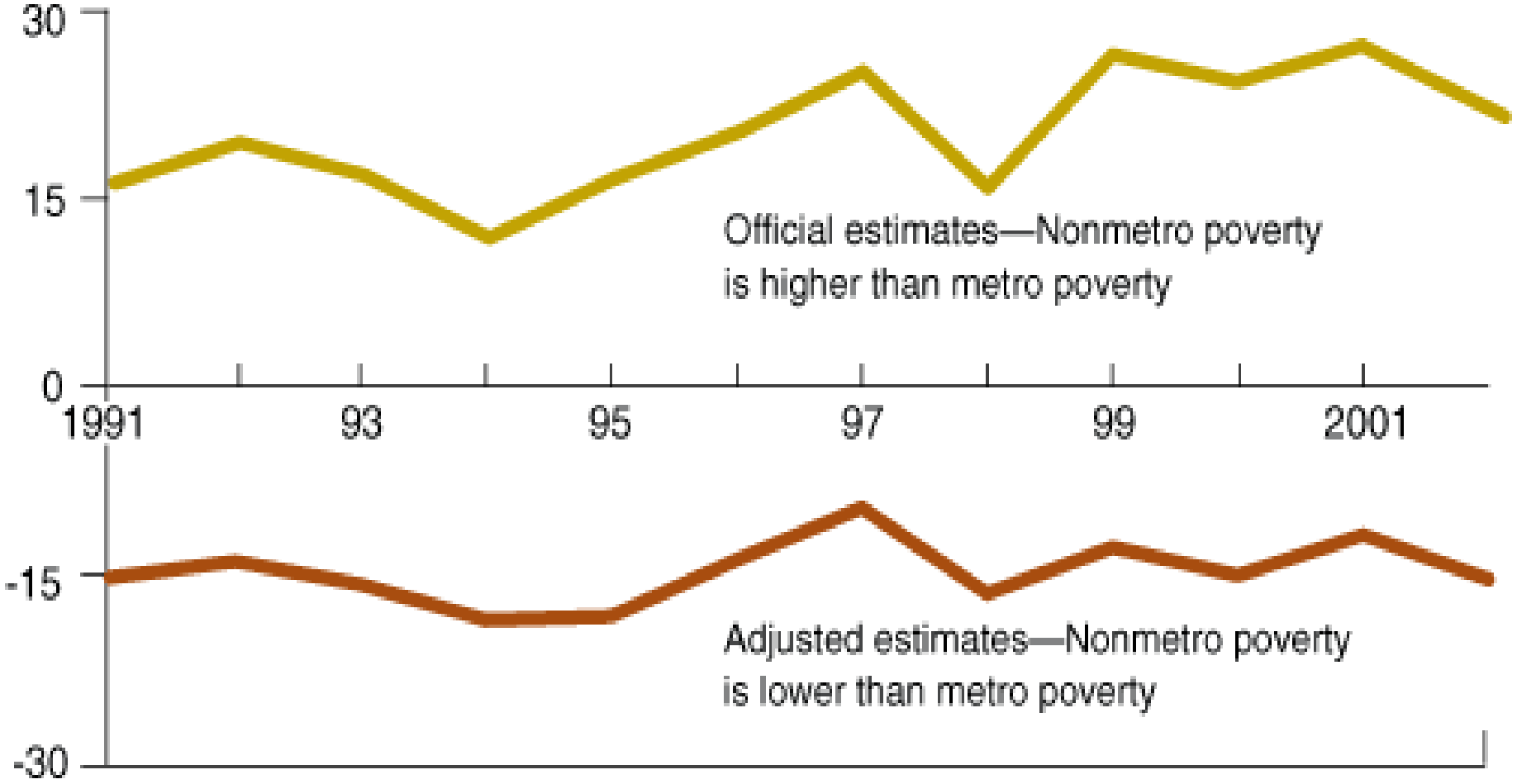


WORLD BANK GROUP

	Headcount, P_0 Measure		Poverty gap, P_1 Measure		Squared Poverty gap, P_2 Measure	
	Actual	FMR adjusted	Actual	FMR adjusted	Actual	FMR adjusted
2001						
Nonmetro	**0.142**	**0.105**	0.063	0.050	0.043	0.036
	(0.004)	**(0.003)**	(0.002)	(0.002)	(0.002)	(0.002)
Metro	**0.111**	**0.120**	0.052	0.055	0.036	0.038
	(0.002)	**(0.002)**	(0.001)	(0.001)	(0.001)	(0.001)
Nonmetro-Metro	**28%**	**-12%**	21%	-9%	18%	-5%
Difference	**(3.80)**	**(3.01)**	(4.71)	(3.97)	(5.64)	(4.97)

Cost-of-living adjustment reverses poverty rankings

Percent difference between nonmetro and metro poverty rates



Source: Current Population Survey, 1992-2003.

Adjustments: Household Composition

Concepts, Example, Comments

Equivalence scales, some motivation

Issue: Income often times measured at the level of the household (e.g., farming households or nonfarm enterprise households); typically shared across the household.

	One adult	Couple- without children	Couple- with children
			
Income	1000 €	1000 €	1000 €

Do these HHs have the same Standard of living?

1. People have different needs. If thinking about social policy with an emphasis on poorer people, then useful to think about differences in basic needs (e.g., calories).

2. There may be economies of scale within the household. Six people maybe able to meet basic needs on less than twice what it costs a family of three to meet basic needs. Consumption of public goods within the household (shelter, light, radio, water pump, …) and potentially economies in food preparation (fuel, time)

Equivalence scales adjust household income to take into account differences in the size (economies of scale) and/or composition of households (adult-equivalence adjustments) to improve comparability of households.

Commonly used adjustment scales

- **Oxford scale**: 1 to the first household member, 0.7 to each additional adult, 0.5 to each child. Let Nc be number of children an Na be number of adults…

$$AE_{Oxford} = 1 + 0.7 * (N_A - 1) + 0.5 * N_c$$

- **OECD-modified scale**: 1 to the household head, 0.5 to each additional adult, 0.3 to each child

$$AE_{OECD} = 1 + 0.5 * (N_A - 1) + 0.3 * N_c$$

- **National Research Council,** 2 parameter scale: α = adult equivalent (0,1); Θ = economies of scale (0,1)

$$AE_{LSMS} = (N_A + \alpha N_c)^{\theta}$$

- **Single-parameter scale**: N^{θ}
  - constant-elasticity equivalence scale, lies between 0 and 1
  - If θ=0, then every individual is assigned household income
  - If θ=1, then income is assigned on a per person basis.
  - *A common implementation of this scale is* θ=0.5

How does the single-parameter scale handle these households

	One adult	Couple- without children	Couple- with children
			
Income	1000 €	1000 €	1000 €

Do these HHs have the same Standard of living??

- Consider the single-parameter scale
- If θ=0, then 1000, 1000, 1000
- If θ=1, then 1000, 500, 333.3income is assigned on a per person basis.
- *If* θ=0.5, then 1000, 707, 577 (e.g., as used by LIS)

Equivalence Elasticity: How needs change with additional household members

"The needs of a household grow with each additional member but – due to economies of scale in consumption– not in a proportional way. Needs for housing space, electricity, etc. will not be three times as high for a household with three members than for a single person. With the help of equivalence scales each household type in the population is assigned a value in proportion to its needs. The factors commonly taken into account to assign these values are the size of the household and the age of its members (whether they are adults or children). A wide range of equivalence scales exist, many of which are reviewed in Atkinson et al. (1995). Some of the most commonly used scales include: " Source: OECD

Household size	**Equivalence scale**				
	per-capita income	"Oxford" scale ("Old OECD scale")	"OECD-modified" scale	Square root scale	Household income
1 adult	1	1	1	1	1
2 adults	2	1.7	1.5	1.4	1
2 adults, 1 child	3	2.2	1.8	1.7	1
2 adults, 2 children	4	2.7	2.1	2.0	1
2 adults, 3 children	5	3.2	2.4	2.2	1
Elasticity[1]	1	0.73	0.53	0.50	0

[1] Using household size as the determinant, equivalence scales can be expressed through an "equivalence elasticity", i.e. the power by which economic needs change with household size. The equivalence elasticity can range from 0 (when unadjusted household disposable income is taken as the income measure) to 1 (when per capita household income is used). The smaller the value for this elasticity, the higher the economies of scale in consumption.

Equivalence scales, focus on Economies of scale

- Assume that ρ is a proportion of household expenditure on purely *private goods* and (1- ρ) is allocated to *public goods.*
- *If* ρ=1, then θ is larger, closer to 1*; If* ρ=0, then θ is smaller, closer to *0*
- *Discuss*
- One way to think about this:

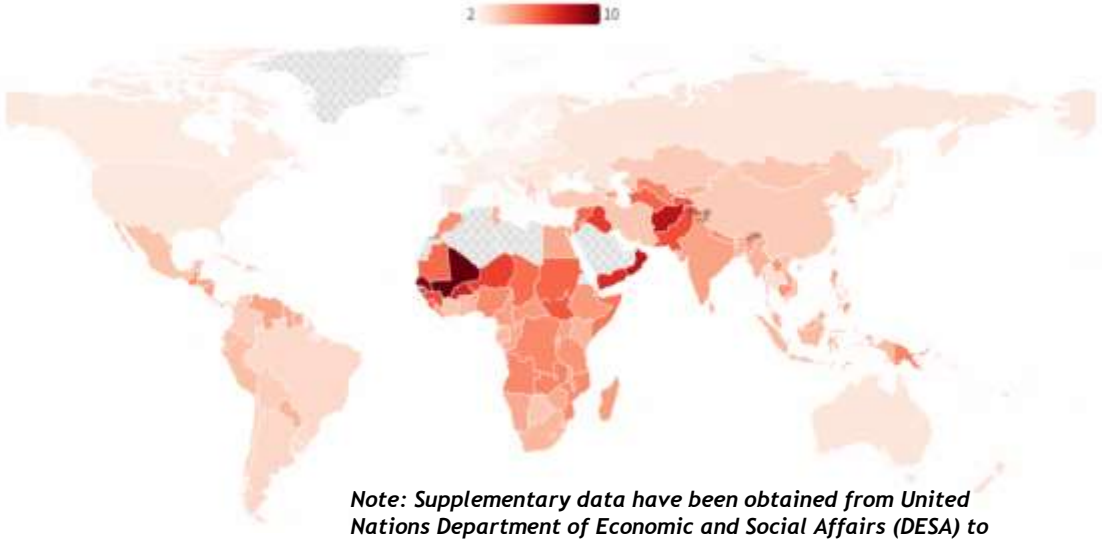
$$x^* = \frac{x}{n^\theta} = \rho\frac{x}{n} + (1-\rho)x$$

- Solving

$$\theta = \frac{-\ln\left(1-\rho+\frac{\rho}{n}\right)}{\ln(n)}$$

- Consider a country with average household size of 4 and on average households spend two-thirds of their income on food (i.e., let's assume ρ =0.66), then $\theta = 0.5$

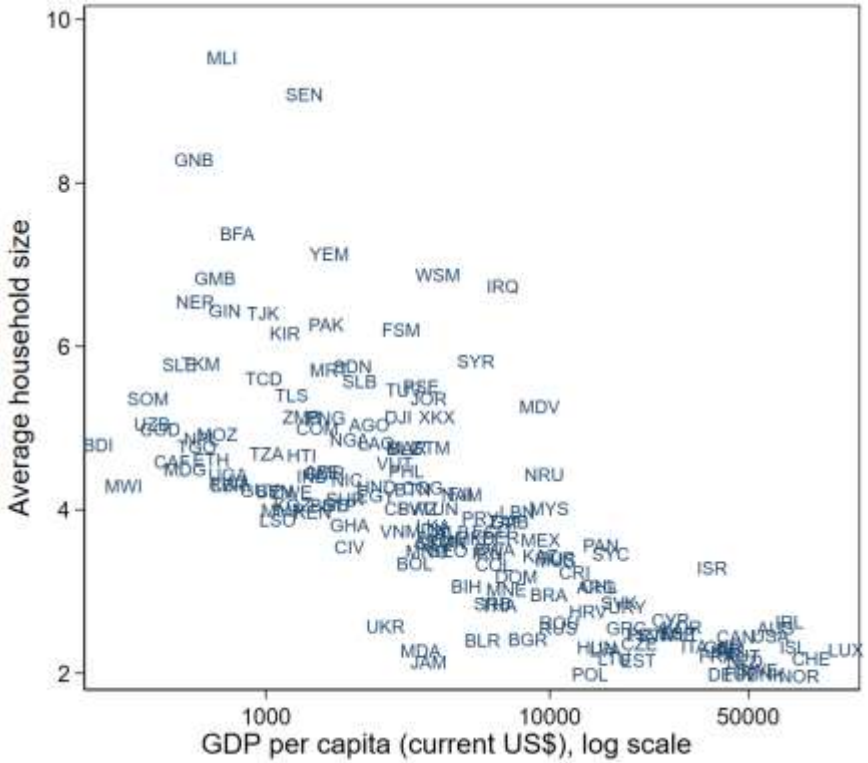
1. Household size varies across countries and regions.



Note: Supplementary data have been obtained from United Nations Department of Economic and Social Affairs (DESA) to improve coverage.

Regions (1)	Household size (2)	Year (3)	Households (in millions) (4)	Countries (5)
Sub-Saharan Africa	4.7	2014.7	207	46
Middle East & North Africa	4.4	2013.5	71	11
South Asia	4.5	2014.8	366	7
East Asia & Pacific	3.8	2013.9	169	19
World	**3.6**	**2014.8**	**1578**	**162**
Latin America & Caribbean	3.3	2013.4	176	22
Europe & Central Asia	2.8	2015.2	171	30
Other High Income	2.4	2016.6	418	27

2. Household size is correlated with income.



***Note**: Household size data are computed using survey data in Global Monitoring Database (GMD) and Luxembourg Income Study (LIS) for circa 2017. GDP data are from the World Development Indicators (WDI), July 2022.*

3. Household size is evolving differently across countries and regions.

Average household size: compare Nigeria and India

Country	1990	1992	2015	Change
Nigeria	5.39		4.90	-9%
India		5.70	4.57	-20%

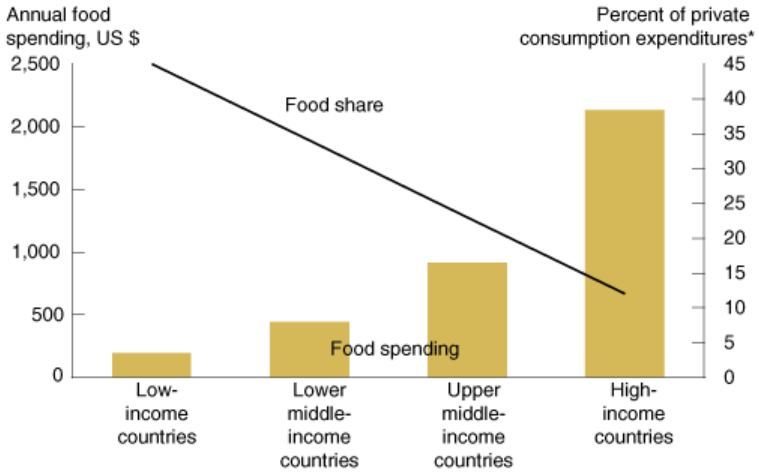
Source: United Nations - Department of Economic and Social Affairs (UN-DESA)

4. Food shares differ by income and shrinking over time

Food shares: compare low-income and high income

Income group	2005	2017
Low-income countries	48.5%	37%
High-income countries	20.4%	5.7%
Change (pp)	28.1	31.3

Source: ICP 2005, 2017
Note: The figures represent the shares of food, beverages and tobacco in GDP in 2005 or shares of food and nonalcholic beverages in GDP in 2017.



Note: * a proxy for income.
Source: Country-level data from Euromonitor International (2006) and country grouping based on 2005 World Development Indicators from the World Bank. U.S. data from U.S. Bureau of Labor Statistics, Consumer Expenditure Survey (2004-2005) for four-person households.

Recall, as ρ gets smaller, then very hard to justify a large θ

Adjusting for economies of scale – a global example

- The per-capita method across countries is hard to justify.
  1. Household size varies across countries and regions.
  2. Household size is correlated with income.
  3. Household size is evolving quite differently across countries and regions.
  4. Food shares are declining due to economic growth.
  5. Food shares differ by income and region.

The World Bank is often criticized because their poverty line is in terms of a given number of dollars a day per person and hence allows for no economies of scale...

But in all cases of poverty measurement, except for the World Bank, everyone agrees that a household-size adjustment is necessary.

Smeeding, T. M. (2016), Poverty measurement. In: Brady, D. and Burton, L. M. (eds.) *The Oxford Handbook of the Social Science of Poverty.* Oxford: Oxford University Press.

Global Example: Questions we aim to shed light on

1. How would the profile of global poverty change if one accounted for economies of scale by deflating household welfare with the square root of household size?
   i. Net changes in regional prevalence rates
   ii. Net changes in country prevalence rates
   iii. How many people switch status (poor ⇔ not poor)

2. Is there any evidence that either per capita or square root allocations does better in terms of identifying the poor?

Results for 2017 - $1.9 (per capita) vs. $4.47 (root N)

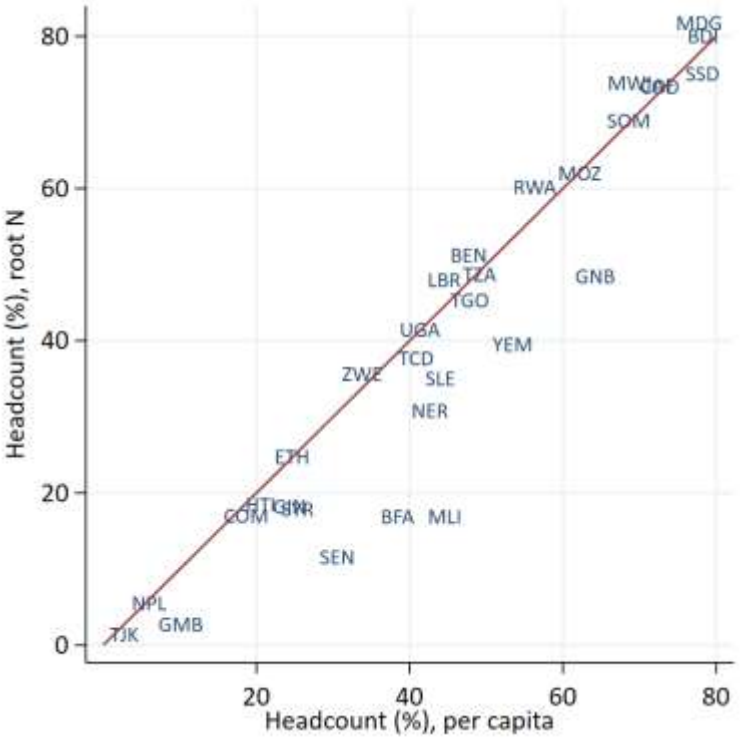
Region	Per capita poverty rate (%) at $1.90 (1)	Root N poverty rate (%) at $4.47 (2)	Change in poverty (pp) (3)	Millions of per capita poor (4)	Millions of root N poor (5)	Change in millions of poor (6)	Absolute deviations in millions of poor (7)	HH size, per capita poor (8)	HH size, root N poor (9)	HH size (10)	Countries (11)
Sub-Saharan Africa	41.28	38.97	-2.31	432	407	-24.2	34.9	6.2	4.9	4.7	46
Middle East & North Africa	7.05	5.26	-1.80	23	17	-5.9	6.3	6.6	4.2	4.4	11
Europe and Central Asia	1.30	1.25	-0.04	6	6	-0.2	1.2	5.2	2.6	2.8	30
World	**11.63**	**11.63**	**0.00**	**682**	**682**	**0.0**	**86.4**	**4.5**	**3.0**	**3.6**	**162**
Other High Income	0.69	0.79	0.10	7	8	1.0	1.0	1.3	1.2	2.4	27
Latin America & Caribbean	3.77	4.41	0.64	22	26	3.7	4.5	4.1	3.2	3.3	22
South Asia	9.62	10.57	0.95	169	186	16.6	27.1	5.9	3.8	4.5	7
East Asia & Pacific	3.63	5.02	1.39	23	32	8.9	11.4	5.3	3.1	3.8	19

Reclassifications moving from per capita to root N rule

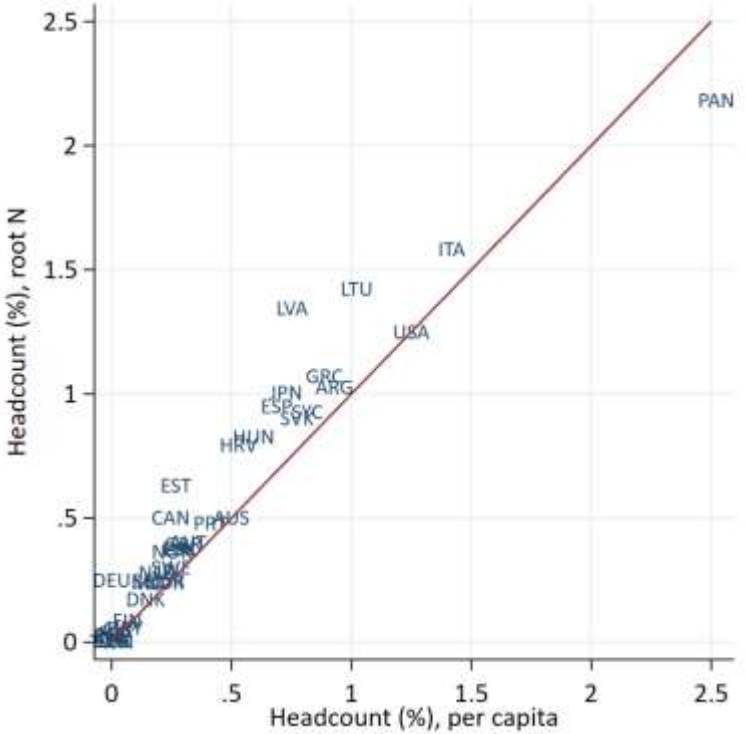
Region	Per capita poverty rate (%) at $1.90 (1)	Root N poverty rate (%) at $4.47 (2)	Change in poverty (pp) (3)	Change in millions of poor (4)	Not poor under both rules (millions) (5)	Poor by pc rule, RECLASSIFIED as not poor (millions) (6)	Not poor by pc, RECLASSIFIED as poor (millions) (7)	Poor under both rules (millions) (8)	Population (millions) (9)
Sub-Saharan Africa	41.28	38.97	-2.31	-24.2	570	68	44	364	1045
Middle East & North Africa	7.05	5.26	-1.80	-5.9	304	8	2	16	329
Europe and Central Asia	1.30	1.25	-0.04	-0.2	483	1	1	5	491
World	**11.63**	**11.63**	**0.00**	**0.0**	**5048**	**135**	**135**	**547**	**5868**
Other High Income	0.69	0.79	0.10	1.0	1012	0	1	7	1020
Latin America & Caribbean	3.77	4.41	0.64	3.7	555	2	6	20	583
South Asia	9.62	10.57	0.95	16.6	1521	50	66	119	1757
East Asia & Pacific	3.63	5.02	1.39	8.9	602	6	15	17	640

Results - $1.90 (per capita) vs. $4.47 (root N) – country level

Low-income countries, 2017



High-income countries, 2017



Summary of descriptive statistics

- Regional profile of poverty is robust to changing allocation rule.

- The root N poverty measure slightly decreases (increases) extreme poverty in the poorest (richest) countries.

- However, there is substantial reclassification of 270 million people as poor or non-poor (20% + 20%).

Which measure does a better job identifying the poor?

- We cannot answer that question but we aim to answer which measure is more highly correlated with non-monetary indicators that are typically assumed to be indicators of poverty.
- *Conditions for selecting a related indicator of well-being:*
  1. A strong relationship between poverty and the indicator is well-established in the literature.
  2. The indicator is orthogonal to household size.

- Most indicators that satisfy (1) hardly satisfy (2)!

  [covariates of poverty: educational attainment, asset ownership, sanitation, employment, etc.]
- We condition out household size from covariates of poverty and use the residuals as instruments of related indicators of well-being.

Purposive sampling of countries & covariates of poverty

- Select at least one country from each region, with two countries each from Sub-Saharan Africa and South Asia, the regions with the highest prevalence of poverty.
- Select countries to maximize the share of the global population covered.
- Select variables presumed to be correlated with poverty.
- Covariates of poverty should have sufficient coverage within and across countries.

Summary statistics on covariates of poverty

Category	Nigeria 2018	Mali 2009	India 2011	Pakistan 2018	Tajikistan 2015	Indonesia 2017	Yemen 2014	Colombia 2017
Years of schooling	7.0		5.5	5.4		8.21	6.12	8.22
Asset index	2.51		2.12		1.77	2.21	2.68	
Asset ownership: computer or landline		0.04		0.14				0.43
Literacy	0.72	0.35	0.68	0.58		0.96	0.71	0.93
Not employed in the agricultural sector	0.92			0.70		0.65		0.80
Access to electricity	0.64	0.22	0.80	0.91	0.98	0.98	0.65	0.98
Piped drinking water	0.03	0.64		0.93	0.46	0.11	0.48	0.98
Improved sanitation	0.58	0.22		0.70	0.96	0.76	0.59	0.90

Method

Estimate a vector of residuals (e) from the following specification:

$$Y = \beta_0 + \beta_1 N + e \qquad (1)$$

where Y is a vector of years of schooling [or asset indices] of household heads

N is a vector of household size

The following conditions must hold:

- $E[Ye] \neq 0$, $E[YP] \neq 0$, hence $E[Pe] \neq 0$ [select a y that can satisfy these conditions]
- $E[Ne] = 0$ [orthogonality condition]

Let P be the probability of being (i) per capita poor, or (ii) root N poor.

Create q quantiles from the vector of residuals (e).

Examine whether e, the residual, is more correlated with being root N poor or per capita poor.

Identifying the extreme poor based on covariates of poverty

Which measure, per-capita or square root of household size, is more highly correlated with typical covariates of poverty?

(Pooled data across countries with each country weighted by millions of poor)

Category **(1)**	**Per capita poor only** **(2)**	**Root N poor only** **(3)**	**Diff. p-value** **(4)**	**Per capita poor** **(5)**	**Root N poor** **(6)**	**Diff. p-value** **(7)**	**Obs** **(8)**
Years of schooling	-0.007***	-0.216***	0	-0.257***	-0.465***	0	683,533
Asset index	-0.052***	-0.185***	0	-0.529***	-0.661***	0	431,436
Asset ownership	-0.010***	-0.053***	0	-0.154***	-0.197***	0	264,680
Literacy	-0.012***	-0.085***	0	-0.155***	-0.227***	0	694,463
Not employed in the agricultural sector	-0.009*	-0.079***	0	-0.130***	-0.199***	0	450,521
Access to electricity	-0.015***	-0.082***	0	-0.262***	-0.329***	0	697,574
Piped drinking water	-0.011***	-0.033***	0.0007	-0.143***	-0.164***	0.0007	595,934
Improved sanitation	-0.008***	-0.058***	0	-0.182***	-0.232***	0	592,937

Equivalence scales for income and consumption; *What about wealth?*

- The adjustments are motivated by consumption needs differing across people and households of different size.
- It's clear that consumption is the target measure meant to be adjusted. Income is a good candidate for adjustment as well because the majority of income is converted into consumption. This is particularly true for lower income households.
- What about wealth?
  - No consensus view around this question, but let me propose that the answer will differ depending on what the measure of wealth is being used for, and/or how wealth is most commonly used in a country.
- Consider a few different scenarios, and think about whether adjustment is helpful
  - Relatively poor country with significant volatility in weather and agricultural productivity. For most households, wealth is a savings mechanism to help maintain food security in bad times. DISCUSS.
  - A country where most households hold wealth as insurance against adverse shocks, but largely do not draw down on this wealth. DISCUSS.
  - A country where wealth is primarily a vehicle for maintaining inter-generational economic status. Wealth is primarily passed down to children. DISCUSS.

Tangent: Asset Indices as Wealth Proxy

Concepts, Concerns

Asset Indices: Presentation outline

Movitation: Asset indices, frequently constructed with principal component analysis (PCA), are treated as measures of wealth. First some background, then some comments on when these are, and are not, useful.

I. PCA: concept & method overview
II. Application: Nigeria (2018/19)
III. Comments and some literature

PCA concept & motivation

- PCA is a statistical technique used for reducing the dimensionality of data sets with the goal to maximize interpretability while minimizing information loss.

- Using PCA is a pragmatic response to a data constraint problem.
  - Lack of wealth data particularly in developing countries creates a demand for the construction of an asset index.
  - Growth in household survey data in developing countries, with multiple indicators of household assets

Feature 1	Feature 2
4	2
6	3
13	6
...	...

Note high level of collinearity.

Much, but not all, of the variation can be explained By recognizing feature 1 co-moves w feature 2 by a factor of 2

PCA is all about leveraging correlations

PCA methodology

- Construct a data set of p numeric variables (e.g., household assets), for each n households. *The p variables need to be correlated, if orthogonal, PCA has no value added. Furthermore, a useful p variable will have variation. If no variation in p, then it cannot help to rank households.*
- The data values define p n-dimensional vectors $\boldsymbol{x_1}, \ldots, \boldsymbol{x_p}$, or equivalently, an $n \times p$ data matrix $\boldsymbol{X}$.
- The objective is to obtain a linear combination of the columns of $\boldsymbol{X}$ that explains as much of the co-variance as possible.
- Such linear combinations are given by $\sum_{j=1}^{p} a_j \boldsymbol{x}_j = \boldsymbol{Xa}$, where $\boldsymbol{a}$ is a vector of constants $a_1, a_2, \ldots. a_p$.
- The variance of any such linear combination is given by $var(\boldsymbol{Xa}) = \boldsymbol{a'Sa}$, where $\boldsymbol{S}$ is the sample covariance matrix associated with the data set.

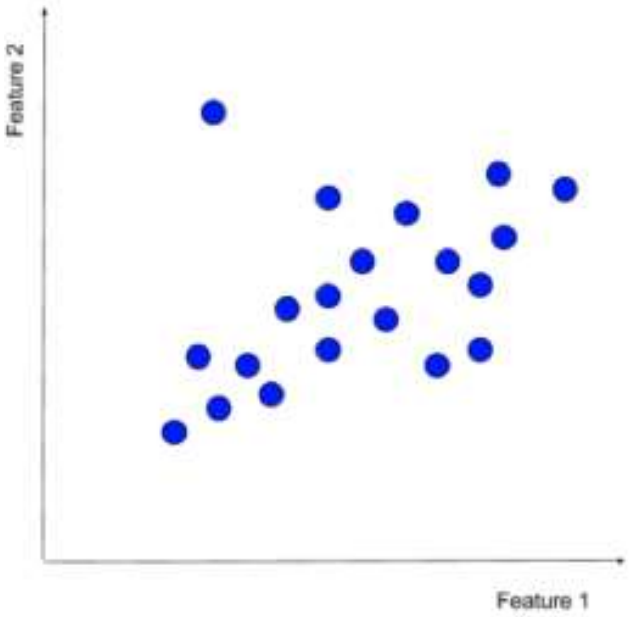
PCA methodology: the eigenproblem

- Identify the linear combination of p's that explains the covariance. If you recall your linear algebra, this can be stated more formally …
- Equivalently, obtain a p-dimensional vector $\boldsymbol{a}$ which maximizes the quadratic form $\boldsymbol{a'Sa}$.
- For a well-defined solution, impose the restriction $\boldsymbol{a'a = 1}$.

$$\max \quad \boldsymbol{a'Sa} \quad \text{subject to} \quad \boldsymbol{a'a = 1}$$

- Maximize $\boldsymbol{a'Sa} - \lambda(\boldsymbol{a'a} - 1)$, where λ is a Lagrange multiplier. Note that S, the co-variance matrix of the p's, is at the heart of solving for the principal components.
- Result is given as: $\boldsymbol{Sa} - \lambda\boldsymbol{a} = \boldsymbol{0} \iff \boldsymbol{Sa} = \lambda\boldsymbol{a}$, where $\boldsymbol{a}$ is a unit-form eigenvector, and λ the corresponding eigenvalue of the covariance matrix $\boldsymbol{S}$.

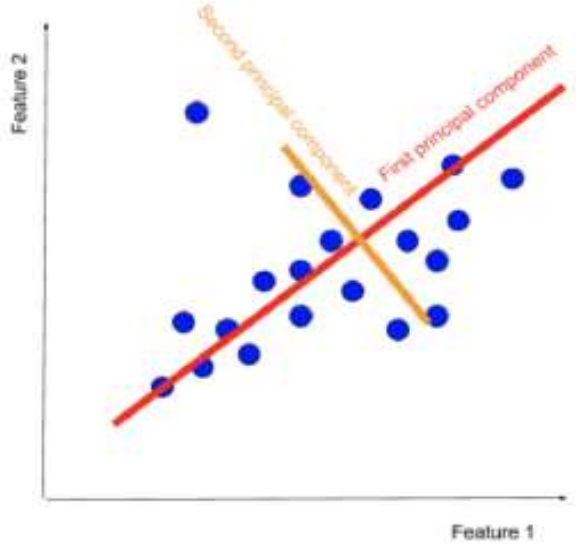
PCA methodology: simple illustration with two dimensions



Note the covariance of the two features.

Estimate/draw a line that explains most of the covariance – the first principal component

Now construct a second line, orthogonal to the first, that explains as much remaining covariance



PCA methodology

- By definition, there are p orthogonal eigenvalues, known as **principal components**.

- The largest eigenvalue, or the first principal component, is the latent underlying factor that explains the maximum variance (and covariance) in the asset variables.

- It is intuitive to **assume** that the first principal component is household long-run wealth. From a statistical view, it's really just a collapsed way of ranking households in terms of asset holdings.

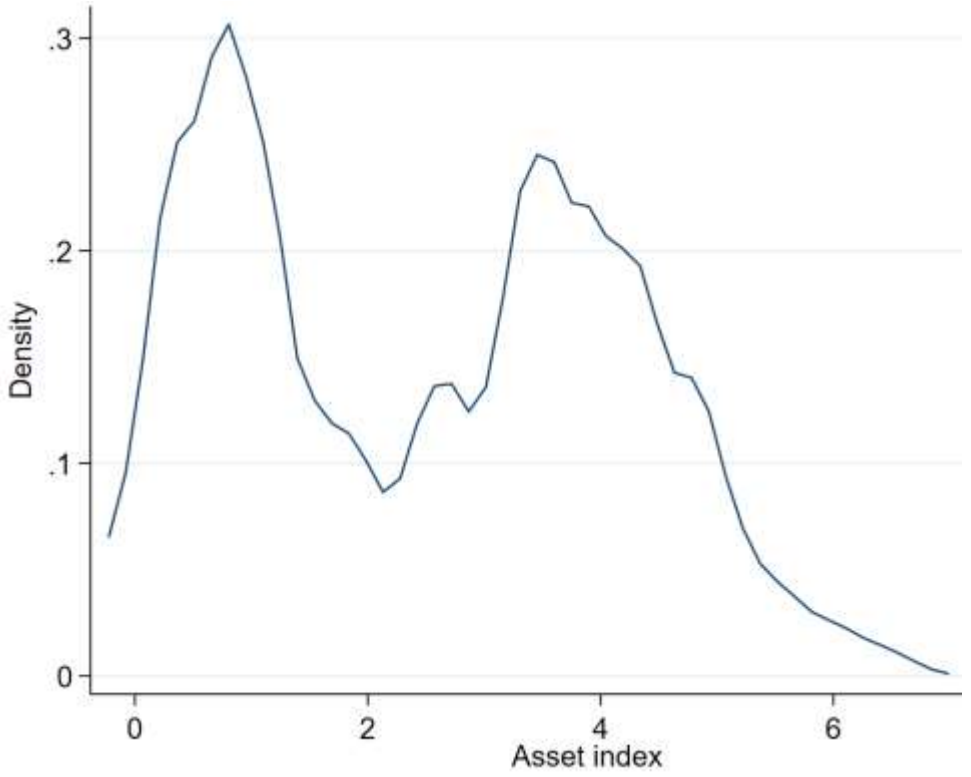
- The asset index performs well in ranking households, in empirical tests conducted with data sets having both expenditure and wealth data (Filmer and Pritchett, 2001). They show the multi-dimensional index is well correlated w wealth.

PCA application: Nigeria (2018/2019)

Asset	Does own asset (=1)
TV	0.817
Fan	0.814
Electricity	0.769
Fridge	0.660
Stove	0.649
Floor material	0.525
Car	0.506
Computer	0.451
Phone	0.416
A/C	0.415
Washing machine	0.383
Sewing machine	0.202
Radio	0.196
Bicycle	-0.019

These are the "factor loadings", the **a** vector, of the **first principal component**.

The ownership of TV sets, fan, electricity, fridge, and stove is most indicative of wealth. The ownership of bicycle is an indicator of poverty.



Caveats and concerns

Inputs to asset index ALL need to be in the same units, typically either YES/NO, or normalized counts. Cannot be categorical, cannot be mixture.

Asset indices are scale dependent. They skew towards those variables with greatest variance. So, changing from dollars to cents will greatly affect loading factors, increase the importance of the variable. (Also: Sensitive to outliers)

Loading factors are weights, analogous to prices for consumption, they change over time. Asset index values have no meaning over time.

Asset index units are not comparable across countries.

When are asset indices helpful?

Where focus is on ranking households by wealth status.

Note: not possible to say how wealthy a household is from asset index, but is possible to rank households.

Can only make comparisons over time if loading factors are stable. Is this likely to happen?

Depends on the time gap. 6 months, probably; 6 years, probably not. Think about cell phone ownership in poor countries 10 years ago, and how much signal that had compared to now.

When primary need is to control for wealth but important to understand, these indices do not measure wealth.

References and suggested links

Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *374*(2065), 20150202.

Filmer, D., & Pritchett, L. H. (2001). Estimating wealth effects without expenditure data—or tears: an application to educational enrollments in states of India. *Demography*, *38*(1), 115-132.

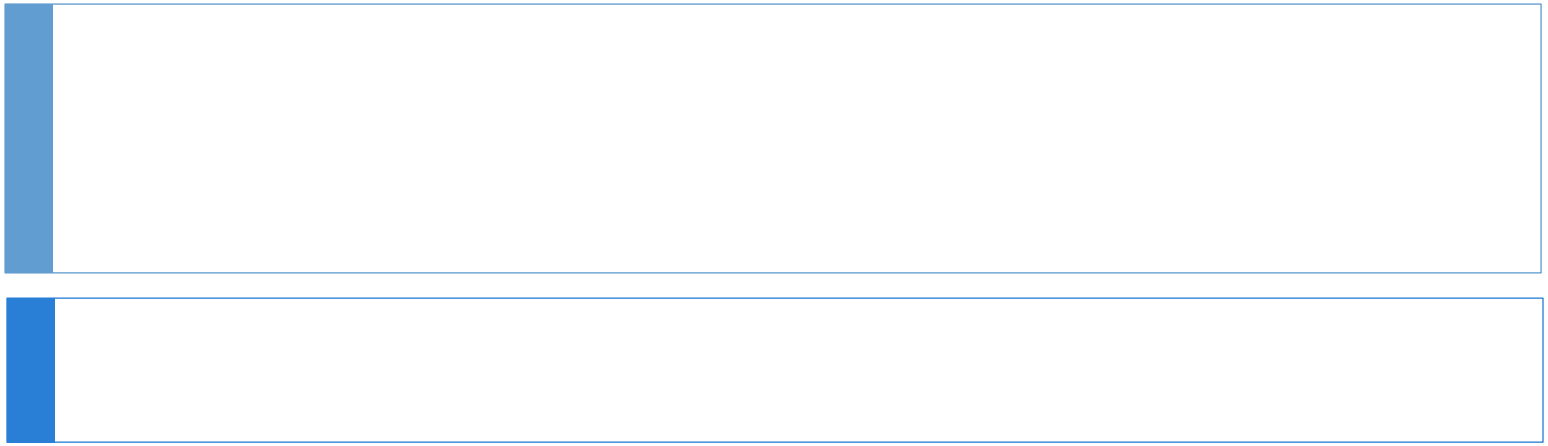
https://ramanlab.wustl.edu/Lectures/Lecture11_PCA.pdf

https://people.duke.edu/~hpgavin/SystemID/References/Gillies-PCA-notes.pdf

https://www.youtube.com/watch?v=BsJJXQ10ayM

https://www.khanacademy.org/math/linear-algebra/alternate-bases/eigen-everything/v/linear-algebra-introduction-to-eigenvalues-and-eigenvectors

Thank You



WORLD BANK GROUP

Typical formula for Paasche and Laspeyres index

- **Paasche** index

$$P_P = \left(\sum_{k=1}^{K} w_{hk} * \frac{P_{Rk}}{P_{hk}} \right)^{-1}$$

- **Laspeyres** index

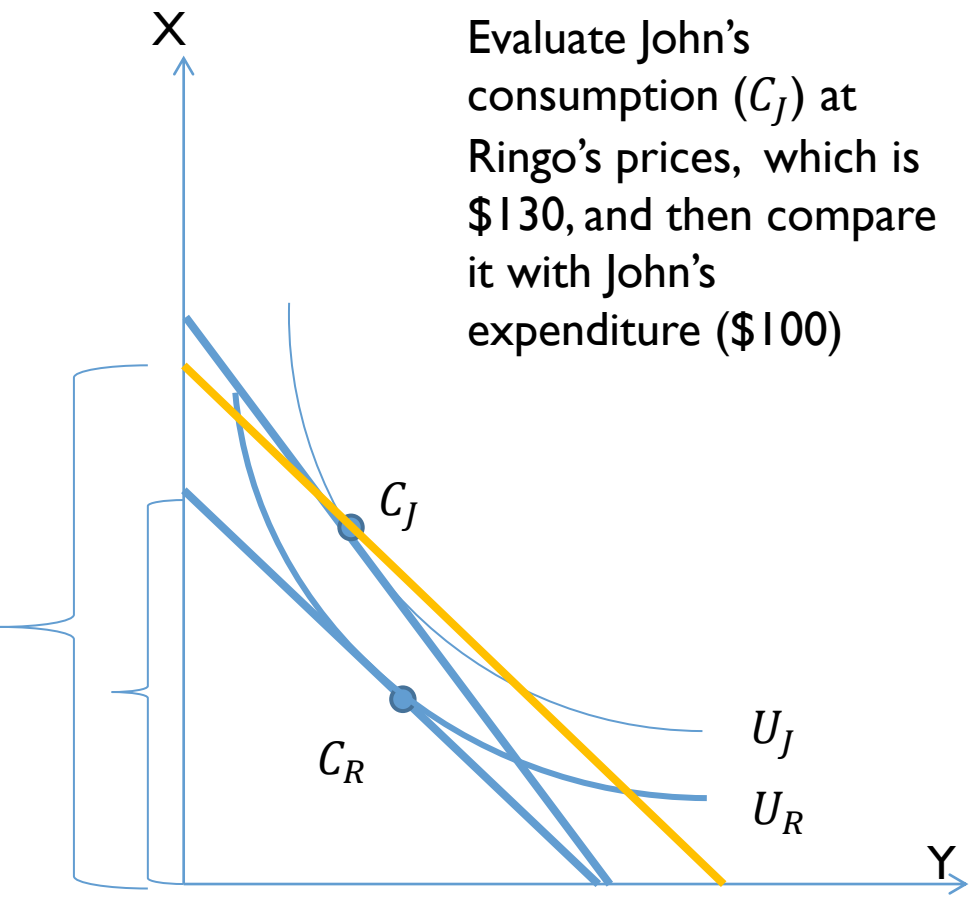
$$P_L = \sum_{k=1}^{K} w_{Rk} * \frac{P_{hk}}{P_{Rk}}$$

Where w_{hk} refers to budget share of item k for household h; R refers to reference

- **Fischer** Ideal index

$$P_F = \sqrt{P_P * P_L}$$

Visual presentation of Laspayres Index



Evaluate John's consumption (C_J) at Ringo's prices, which is \$130, and then compare it with John's expenditure (\$100)