AGRODEP Training Session

"Poverty measurement and analysis"

Anne-Claire Thomas & Philippe Van Kerm
Université Catholique de Louvain (Belgium)
& CEPS/INSTEAD (Luxembourg)

International Food Policy Research Institute (IFPRI), Dakar
April 24–27 2012

Session 1

Measuring poverty: the basic tools

Content

[outline]

Quantifying poverty

Strictly speaking, poverty (at individual- or household-level) is defined as possessing insufficient resources to reach a ‘satisfactory’ living standard

“Satisfactory”?

Quantitative, monetary approach (narrowly defined perhaps – see session on multidimensional approach): assume we can set a minimum level of resources (income, expenditure, in ‘single adult equivalents’) z –the poverty line– under which one can not maintain an acceptable living standard

Focus on the bottom of the distribution

Three questions

Three questions:

- *Who* is poor? Identification
  - (NB: in EU, official statistics refer to being ‘at *risk* of poverty’)
- What is the intensity of poverty of an individual?
- What is the aggregate level of poverty in the population? Aggregation

Identification: the poverty line

A person is in poverty (or ‘at risk of poverty’) if the equivalent income y of the household to which he belongs to is below z

z is the *poverty line*

An important distinction

- ‘absolute’ poverty line
- ‘relative’ poverty line

In any case, fair amount of arbitrariness in this discontinuity

A relative poverty line

In OECD countries, typical approach is to use a ‘relative’ poverty line

Usually, z is determined as a pre-defined fraction of a reference income:

$$z = f\,Y$$

where Y is typically the median income (sometimes the mean income) and where f is ... (what do you think?) ...

A relative poverty line

In OECD countries, typical approach is to use a ‘relative’ poverty line

Usually, z is determined as a pre-defined fraction of a reference income:

$$z = f\,Y$$

where Y is typically the median income (sometimes the mean income) and where f is ... (what do you think?) ... between 0.50 et 0.70 (EU measures use $f = 0.60$)

(NB: the choice of f is largely arbitrary)

An absolute poverty line

z reflects the monetary value of a 'minimal' consumption bundle (e.g., to reach a given calorie intake, e.g., 2,250 calories norm per day)
Typically determined by 'expert groups'

- food intake only?
- allowance for non-food items?

Once determined, the value is updated for price increases over time (the reference bundle is normally held constant)

Variants

- A ‘mixed’ approach: z determined by a relative approach at one point in time... but then updated by variations in prices (not variations in the reference level)
- A combined approach: $z = (z^r)^\theta \times (z^a)^{1-\theta}$
- Subjective determination: use survey questions to assess the poverty line. Respondents are directly asked how much they perceive as required to get along in the society.

In any case, arbitrariness is inevitable: avoid putting to much emphasis on discontinuity at the poverty line and rely on methods robust to choice of poverty line (presented shortly)

Intensity of poverty: the poverty gap

All poor people are not equally poor: the ‘poverty gap’ gives an idea of the intensity of the poverty experience of a person with income y_i

$$g_i = \max\left(\frac{z - y_i}{z}, 0\right)$$

Gap is zero for non poor individuals, and is equal to relative shortfall from the poverty line (expressed as fraction of poverty line) for poor people. (Note that it is independent of the units of the data – help comparisons over time and space)

Aggregation: properties of measures of poverty

Key property of poverty indices is ‘focus’ (which makes it different from inequality or aggregate welfare measures): the measure is not influenced by income of non-poor people

Additional relevant properties:

- ‘monotonicity’: poverty reduced (or not increased) by increase in income of a poor person (and all other poor people unaffected)
- ‘transfer principle’: transfer from poor person to an even poorer person reduces poverty – inequality aversion à la Pigou-Dalton (but focused on the poor)
- (‘transfer sensitivity’: poverty-reducing impact of a poor-to-poorer transfer is bigger if taking place at lower income levels)

(Many other properties can and have been advocated, see Zheng (1997) for a thorough review)

The headcount index

Poverty rate, or headcount index, merely looks at proportion poor:

$$H = \frac{1}{N}\sum_{i=1}^{N} \mathbf{I}(y_i \leq z) = F(z)$$

Sharp discontinuity at z and H is insensitive to the intensity of poverty (it is cost-efficient to reduce poverty according to H by giving transfers to the least poor first, so that they leave poverty with small transfers – not a very attractive feature!)

Poverty gap ratio

The ‘poverty gap ratio’ – the average gap:

$$P = \frac{1}{N}\sum_{i=1}^{N} g_i = \frac{1}{N}\sum_{i=1}^{N} \max\left(\frac{z-y_i}{z}, 0\right)$$

Not to be confused with the average gap among the poor:
$IG = \frac{1}{N\,H}\sum_{i=1}^{N} g_i = \frac{P}{H}$

The Foster-Greer-Thorbecke family

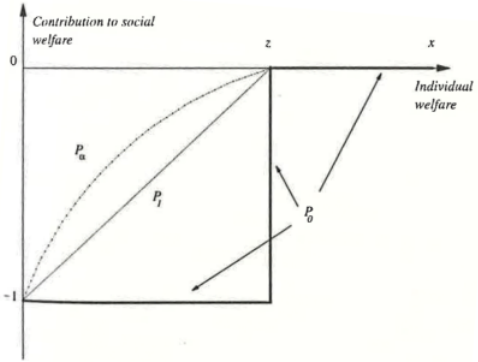
The ‘Foster-Greer-Thorbecke’ (FGT) index is a generalization of the previous two:

$$P_{\alpha}^{\text{FGT}} = \frac{1}{N}\sum_{i=1}^{N} g_i^{\alpha}$$

FGT(0) is the headcount ratio, FGT(1) is the ‘poverty gap ratio’. Taking $\alpha > 1$ assigns greater weight to gap of the poorest (satisfy a transfer principle (unlike $\alpha < 1$!))

The Foster-Greer-Thorbecke family

Poverty contributions for different α parameters (from Deaton (1997, p.145)):



The Watts index

The Watts index (Watts, 1968) is defined as:

$$P^{\text{WATTS}} = \frac{1}{N}\sum_{i=1}^{N} \ln\left(\frac{z}{y_i}\right) \mathbf{I}(y_i \leq z)$$

It is approximately the average proportional shortfall from the poverty line

The Sen index

Sen was 'pioneer' in putting explicit weight on inequality among the poor in an axiomatic framework. The index proposed by Sen (1976) has a different structure:

$$P^{\text{SEN}} = H \times IG \times (1 + G^p)$$

where G^p is the Gini coefficient of poverty gaps income among the poor (formulation from Xu and Osberg (2002))
It can also be written

$$P^{\text{SEN}} = \frac{1}{N} \sum_{i=1}^{q} 2 \frac{q + 0.5 - i}{q} g_i$$

where q is the number of poor, so it is a weighted average poverty gap, with weight linearly decreasing with individual ranks from the poorest to non-poor

The Sen-Shorrocks-Thon index

Variants of the Sen index have been proposed to address some (mild) ‘deficiencies’. E.g., the SST index ‘improves upon’ Sen’s original index by weighting by rank over entire distribution (not only the poor):

$$P^{\mathrm{SST}} = \frac{1}{N}\sum_{i=1}^{N} 2\frac{N + 0.5 - i}{N}\, g_i$$

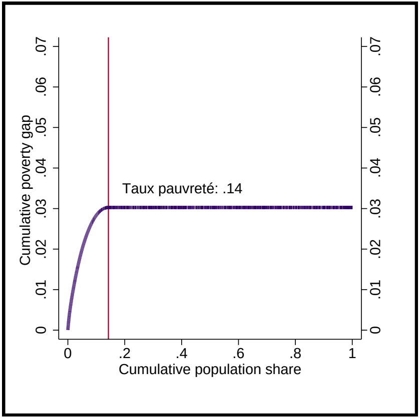
Again a weighted average poverty gap, with weight linearly decreasing with individual ranks, but now from the poorest to richest

$$P^{\mathrm{SST}} = H \times IG \times (1 + G)$$

where G is Gini of poverty gaps in the total population

'Three I's of Poverty' (TIP) curves

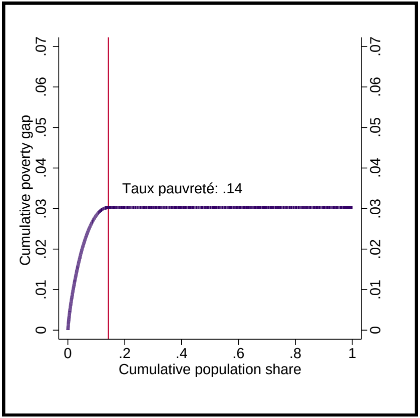
The TIP curve illustrates three I's of poverty: Incidence, Intensity, Inequality



It is a cumulation of (normalized) poverty gaps from biggest to lowest: $TIP(p) = \frac{1}{N}\sum_{i=1}^{Np} g_i$ with $0 \leq p \leq 1$

‘Three I’s of Poverty’ (TIP) curves

The TIP curve illustrates three I’s of poverty: Incidence, Intensity, Inequality



FGT(0) and FGT(1) are directly identified
Curvature shows inequality

[outline]

Estimation with survey data

The various indicators and methods presented so far are easily estimated from survey data (or registers) using formulae given (or obvious sample analogues).
Note however, two issues with sample data:

- Dealing with sampling weights (due tu unequal sampling probabilities)
- Inference (taking survey design into account)

Using sample weights

Incorporating sampling weights is usually straightforward

$$\mu_y = \frac{1}{\sum_{i=1}^{N} w_i} \sum_{i=1}^{N} w_i y_i$$

$$F(y) = \frac{1}{\sum_{i=1}^{N} w_i} \sum_{i=1}^{N} w_i \mathbf{I}(y_i \leq y)$$

($\mathbf{I}(A) = 1$ if A is tru 0 otherwise)

Using sample weights

A weighted covariance (for Gini estimation) is

$$\mathrm{Cov}(x,y) = \left(\frac{\sum_{i=1}^{N} w_i}{(\sum_{i=1}^{N} w_i)^2 - \sum_{i=1}^{N} w_i^2} \right) \sum_{i=1}^{N} w_i (x_i - \mu_x)(y_i - \mu_y)$$

Lorenz curve

$$L(p) = \frac{\sum_{i=1}^{J} w_i y_i}{\sum_{i=1}^{N} w_i y_i}$$

where$J$ is such that $\frac{\sum_{i=1}^{J} w_i}{\sum_{i=1}^{N} w_i} = p$

Two main approaches

Two main approaches for variance estimation, constricition of confidence intervals, tests

- analytic, linearization approaches
- empirical, resampling-based approaches

Variance estimation by linearization

general principles

- θ is the statistic of interest, estimated by $\hat{\theta}$
- A linearization variable Z for θ, is a linear variable ($\hat{Z} = \sum_i w_i z_i$) such that
$$\text{Var}(\hat{Z}) \approx \text{Var}(\hat{\theta})$$
- Once we know z_k, it is easy to estimate $\text{Var}(\hat{Z})$ and therefore $\text{Var}(\hat{\theta})$ (it is the variance of a total – methods are well-known to estimate this with various complex survey design)
- Deville (Survey Methodology, 1999) demonstrates that the 'influence function' (IF) of θ is a valid linearization variable, and gives rules to compute the IF for a variety of statistics. (Other linearization approaches have been used too.)

Linearization variables for poverty measures

Berger & Skinner (App. Statist., 2003) use Deville's method to derive the IF for the low income proportion

- Ignoring estimation of z

$$z_k = \frac{1}{N}\left(\delta\{y_k \leq z\} - \hat{p}\right)$$

- With estimation of $\hat{z} = \alpha \hat{\mathrm{Med}}$

$$z_k = \frac{1}{N}\left(\left(\delta\{y_k \leq \hat{z}\} - \hat{p}\right) - f(\hat{z})\frac{\alpha(\delta\{y_k \leq \hat{\mathrm{Med}}\} - 0.5)}{f(\hat{\mathrm{Med}})}\right)$$

(similar shape for broader class of measures, also if mean is reference income)

Resampling-based inference

Often, textbooks/articles give analytical formulae to estimate standard errors (as per above). But in some non-standard cases the derivations are intractable, do not exist, or you can't find it anywhere...

Resampling-based methods can help

The bootstrap

- Principle: The sample is an ‘clone’ of the population...
- ... you simulate sampling from this population by drawing re-samples (with replacement) ...
- ... compute your estimate in each of these re-samples...
- ... and assess the sampling variability by the variability across all re-samples
- Number of replications needs be high to have accurate estimates (e.g. 1000)
- Leads to consistent estimates of SEs
- Some procedures provide “asymptotic refinements” over asymptotic analytic formulae (better performance in finite samples)
- Caveat: slow (real limitation if point estimation itself time-consuming)

Bootstrap confidence intervals

- Various possibilities can be considered for building CIs
  - If sampling distribution of estimator is normal, plug bootstrap SE into classical CI formula for a normal distribution (default reported in Stata)
  - If you're not sure it is normal, can simply sort your bootstrap estimates and take $\alpha/2$th and $(1-\alpha/2)$th values as CI boundaries (also computed by Stata)
  - Other methods can provide improved precision, but often much more time-consuming, e.g. Bias-corrected and accelerated bootstrap or double-bootstrap

The jackknife

- Another resampling-based method
- ‘deterministic’ resampling (unlike bootstrap)
- remove one obs. and repeat estimation; repeat by removing all obs. in turn
- by combining all the estimates, one can estimate the SE of the orignal estimate
- sometimes results in ‘closed from’ expressions estimable without actually replicating!

Complex design

With any resampling method, the survey design still needs to be taken into account!
In particular, it is essential to

- resample 'PSUs' together
- resample with sampling strata

Current practice typically with 'delete-one-psu jackknife' or 'block bootstrap' (resampling PSU's with replacement) –usually OK if large number of PSUs per stratum–

but resampling inference with complex survey still a subject of ongoing research

[outline]

Dominance

Dominance checks are used to make ‘robust’ comparisons of poverty in two distributions

- robust to choice of a particular index, and/or
- robust to choice of a particular poverty line, and/or
- robust to choice of a particular equivalence scale

Robustness has a cost: ordering is only ‘partial’. Sometimes, one may not be conclusive based on dominance checks only

Classes of poverty measures

Three classes of poverty measures are considered here:

The first two are denoted as in Lambert (2001) $\mathbf{P_A}$ and $\mathbf{P_A^*}$. They are of the generic additive form:

$$P = \int_0^z \theta(y, z) f(y) dy$$

with $\theta(y, z)$ continuous and decreasing in y being individual poverty 'contributions'. $\mathbf{P_A^*}$ is a sub-class of $\mathbf{P_A}$ in which $\theta(y, z)$ is also convex in y.

$\mathbf{P_A}$ includes, e.g., the FGT indices (not strictly for the headcount index), Watts index but not Sen's index. $\mathbf{P_A^*}$ excludes the headcount or an FGT index with $\alpha < 1$.

Classes of poverty measures (ctd.)

The third class is a further restricted subset defined over (normalized) poverty gaps, denoted as in Lambert (2001) $\mathbf{P_{NG}}$. It is of the form:

$$P = \int_0^z \Phi(g) f(y) dy$$

where g is the normalized poverty gap $\frac{z-y}{z}$ and $\Phi(0) = 0$, $\Phi'(g) > 0$, $\Phi''(g) > 0$

This class obviously includes the FGT index (provided $\alpha \geq 1$), but also Watts' index (with $\Phi(g) = -\ln(1-g)$). The main difference is now that the scale of units is irrelevant.

First-order poverty dominance

Consider two distributions A and B sharing a common poverty line z. Atkinson (1987) shows that if the CDF of A is below the CDF of B up to income level Z, then poverty is lower in A according to any poverty index in class $\mathbf{P_A}$ and for any *common* poverty line $z \leq Z$.

Second-order poverty dominance

Consider again two distributions A and B sharing a common poverty line z. Atkinson (1987) also shows that if the integral of the CDF of A is below the integral of the CDF of B ($\int_0^y F^A(s)ds \leq \int_0^y F^B(s)ds$) up to income level Z, then poverty is lower in A according to any poverty index in class $\mathbf{P}^*_{\mathbf{A}}$ and for any *common* poverty line $z \leq Z$.

The plot of $\int_0^y F^A(s)ds$ against y is sometimes called the *poverty deficit curve* and can be shown to be equivalent to a plot of $\int_0^y (y - s)f(s)ds$ against y

TIP dominance

The common poverty line case

Jenkins and Lambert (1997) show that if a TIP curve for A is everywhere below the TIP curve for B (for a common poverty line Z), then poverty is lower in A than in B for any index in $\mathbf{P_{NG}}$ *and* for any common poverty line $z \leq Z$

TIP dominance

The separate poverty lines case

Consider now a situation in which distribution A has poverty line Z_A and distribution B has poverty line Z_B – poverty gaps are defined in the two situations with the respective poverty lines (not a common line anymore).

Jenkins and Lambert (1997) show that if the TIP curve for A is everywhere below the TIP curve for B (for the separate povety lines Z_A and Z_B), then poverty is lower in A than in B for any index in $\mathbf{P_{NG}}$ *and* for any pair of poverty lines kZ_A and kZ_B (with $k \in (0, 1)$).

(Some results exist when TIP curves cross once – see Jenkins and Labmert (1997).)

Sequential poverty dominance

Remember the problem of comparing poverty for households with different needs. Typically this is addressed by applying a conversion function to all incomes to convert to a common reference household type (the equivalence scales)

Potential issue about specific form of equivalence scale: sequential poverty dominance attempts to make robust comparisons

Consider total household income (not equivalized). Assume K household types can be ranked in increasing order of 'needs'. A different poverty line Z_k is chosen for each type: the most needy group will have the higher line: $Z_1 \geq Z_2 ... \geq Z_K$.

Sequential poverty dominance (ctd.)

Assume that we are interested in indices of the type $\mathbf{P}^{*}_{\mathbf{A}}$. For simplicity, assume the same index is used for each household type (although that can be relaxed).

Let F_k be the CDF of income for type k households and p_k be the proportion of these households in the population. Define

$$T_j(z) = \sum_{k=1}^{j} \int_0^z p_k F_k(y) dy$$

($T_j(z)$ is the poverty deficit curve for the sub-population of pooled households of type $k \leq j$ (at least as needy as type j).)

Sequential poverty dominance (ctd.)

Chambaz and Maurin (1998) show that if *for each j* the curve $T_j^A(z)$ for distribution A is below $T_j^B(z)$ for distribution B *for all* $z \leq Z_j$ (using A and B shares of the population types), then poverty is lower in A than in B for any index of type $\mathbf{P}^*_{\mathbf{A}}$, (and for any sets of poverty lines $z_j \leq Z_j$ (provided ordering of lines by need is preserved)).

This is a sequential test: start by checking (second order) poverty dominance for the neediest group up to its poverty line Z_1, then add the second neediest group and check dominance up to Z_2 ($Z_2 \leq Z_1$), etc. up to the end.

[outline]

Measures of poverty

Estimation from survey data

Robust poverty comparisons: dominance

Poverty profiles and decompositions
- Poverty profiles – decompositions by sub-group
- Decomposing poverty change into growth and redistribution
- SST index decomposition

Descriptive regressions

Sub-group decompositions of additive measures

Some of the aggregate measures of poverty can be decomposed into sub-group contributions ('poverty profiles'): e.g., to show contributions by region of residence, by family composition, etc.

This is the attractive feature of additive measures, such as FGT or Watts indices.

Partition population into M sub-groups, and denote s^m the share of population in each of the M sub-groups.
Aggregate poverty is

$$P_\alpha = \sum_{m=1}^{M} s^m \, P_\alpha^m$$

Sub-group decompositions

Analysis of profiles typically consists in reporting

- The poverty level in subgroup m: P_α^m
- The relative poverty risk in subgroup m: $\frac{P_\alpha^m}{P_\alpha}$
- The contribution of subgroup m to total poverty: $\frac{s^m P_\alpha^m}{P_\alpha}$

Sub-group decompositions

- By the same logic, partitions can be nested.
- Contributions are index-specific
- (Distinction between subgroup decomposability and subgroup consistency – Sen's index is none of the two)
- (No role for 'between-group poverty')

Sub-group decompositions for changing poverty

Subgroup decompositions can be used to identify what contributed to *change* in poverty over time: contribution of changes in subgroup population shares vs. changes in subgroup poverty rates

Compute what would poverty be if subgroup shares had remained as in year 1 but with subgroup poverty rates of year 2. Contrast this with actual poverty in year two. Similarly by fixing subgroup poverty rates, etc.

Index number issue (see *supra*)

Growth-redistribution decompositions

How much of a change in poverty can be attributed to changes in the *(relative) distribution* of living standards and how much can be attributed to change in the *average levels* of living standards?

Decompose change in poverty index using counterfactual constructs. One decomposition is:

$$P_2 - P_1 = G(1,2,r) + D(1,2,r) + R(1,2,r)$$

G is a growth contribution, D is a dispersion contribution, and R is a residual

Approach independent on index used

Growth-redistribution decompositions

Index number problem

Let $\tilde{P}_t(r,s)$ be the poverty at time t that would be observed with the level of income observed at time r and dispersion observed at time s. Obviously: $P_t = \tilde{P}_t(t,t)$.

- $G(1,2,r) = \tilde{P}_2(2,r) - \tilde{P}_1(1,r)$
- $D(1,2,r) = \tilde{P}_2(r,2) - \tilde{P}_1(r,1)$

$\tilde{P}_t(r,s)$ can be constructed by computing poverty of the vector of incomes composed of $\mu_r \times \frac{F_s^{-1}(F_t(y_i^t))}{\mu^s}$

Growth-redistribution decompositions

Index number problem

Results will vary for different choice of r (that is for different reference over which the change in means or dispersion is assessed). This causes the residual.

A decompostion without residual is $P_2 - P_1 = G(1,2,1) + D(1,2,2)$. Or $P_2 - P_1 = G(1,2,2) + D(1,2,1)$. This is the index number problem.

One solution is to average contributions over the two possible choices (simplest 'Shapley' value decomposition)

Growth-redistribution decompositions

Combining methods?

One could conceivably combine growth-dispersion decompositions with subgroup decompositions: identify contribution of change in subgroup shares, change in subgroup income levels, change in inequality ‘within’ and ‘between’ groups

Index number issue:

- sequential approach vs. marginal approach
- Shapley value?

(How would you do this? Take this as challenge.)

Multiplicative decomposition of change in SST index

The SST index is not subgroup decomposable, but its change has a particular multiplicative decomposability since it can be written has product of poverty rate (incidence), income gap ratio (intensity) and an inequality component:

$$\frac{dP^{\text{SST}}}{P^{\text{SST}}} = \frac{dH}{H} + \frac{dIG}{IG} + \frac{d(1+G)}{(1+G)}$$

(The last term is apparently typically very small. See Xu and Osberg (2002))

[outline]

Binary choice models

A ubiquitous way to assess relationships between poverty and 'covariates' is using regression models.

'Descriptive regressions': no attempt made to identify causal effects, but identify magnitude of associations (see another session of impact assessment for 'causality')

Binary data, so use of binary choice models, typically probit or logit:

$$\Pr(y_i \leq z | X_i) = \Phi(X_i \beta)$$

where $\Phi()$ is standard normal CDF (for a probit model) and β are regression coefficients on covariates X

Marginal effects or discrete effects

Because of non-linearity of the model, often easier to look at marginal effects: $\beta\phi(X_i\beta)$

But marginal effects vary with X:

- Evaluate marginal effect at the mean: $\beta\phi(\bar{X}\beta)$
- Evaluate marginal effect at other relevant covariate configuration: $\beta\phi(X^R\beta)$
- Average marginal effect: $\frac{1}{N}\sum_i^N \beta\phi(X_i\beta)$

For discrete covariates: look at discrete changes in probability

Alternative regression approaches

- Censored regression (tobit) can be used to model poverty gaps
- Model income distributions and draw implications for any poverty index:
  - e.g. using flexible parametric distributions, e.g. Singh-Maddala (Biewen and Jenkins 2005)
    See last session for more elaborate approaches