

# Computing poverty measures with survey data

Philippe Van Kerm

CEPS/INSTEAD, Luxembourg & ISER, University of Essex  
`philippe.vankerm@ceps.lu`

7th German Stata Users Group Meeting – IZA Bonn, June 26 2009

# Measures of poverty

- ▶ Poverty line:  $z$ 
  - ▶ absolute approach ( $z = \zeta$ ) vs. relative approach ( $z = \beta Y^R$ )  
(where  $Y^R$  is an income reference, typically the mean or median)
- ▶ Most classic measure: The low income proportion  $p = F(z)$   
(headcount ratio)
- ▶ Foster-Greer-Thorbecke family:
  - ▶  $FGT(\alpha) = \int (1 - \frac{x}{z})^\alpha f(x) dx$  ( $\alpha \geq 0$ )
  - ▶ (sensitive to the size and inequality of income shortfall)
- ▶ Watts, Sen-Shorrocks-Thon, median income shortfall (Laeken indicator), Chakravarty, Clark et al. measures, ...

# Some popular user written commands

(selected from recent `findit poverty`)

- ▶ STB-48 sg108 . Computing poverty indices (`help poverty` if installed), P. Van Kerm 3/99 pp.29–33
- ▶ STB-51 sg117 . Robust std errors for Foster-Greer-Thorbecke poverty indices (`help sepov` if installed), D. Jolliffe and A. Semykina 9/99 pp.34–36
- ▶ `povdeco` from <http://fmwww.bc.edu/RePEc/bocode/p> (by S. Jenkins)

# DIY

(or Do I really need a special command for this?)

```
summarize y
scalar z = r(mean)/2
gen byte poor = (y<=z)
mean poor

gen fgt2 = poor * (1-y/z)^ 2
mean fgt2
```

And you get standard error estimates! ...

... and it works with survey data! (svy: mean poor fgt2) ... or does it?

## DIY (ctd.)

(or Do I really need a special command for this?)

First, some measures are not as simple as a mean (typically the Sen-Shorrocks-Thon (SST) index).

Second, in a 'relative approach',  $z$  is *estimated* from the data and estimation of  $z$  impacts on standard errors in ways not accounted for by mean. (Bootstrap? Jackknife?)

# Variance estimation by linearization

## general principles

- ▶  $\theta$  is the statistic of interest, estimated by  $\hat{\theta}$
- ▶ A linearization variable  $Z$  for  $\theta$ , is a linear variable ( $\hat{Z} = \sum_i w_i z_i$ ) such that

$$\text{Var}(\hat{Z}) \approx \text{Var}(\hat{\theta})$$

- ▶ Deville (Survey Methodology, 1999) demonstrates that the 'influence function' (IF) of  $\theta$  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  $z$

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

- ▶ With estimation of  $\hat{z} = \alpha \hat{M}ed$

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

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

# Estimation in Stata

## Option 1 – DIY

Option 1: do steps in .do or .ado file

- ▶ estimate  $\hat{\theta}$  (e.g., `poverty y [aw=w] , ...` )
- ▶ compute the relevant variable  $\hat{Z}$  (e.g., `gen z = ...` )
- ▶ estimate the standard errors with `total` (e.g., `svy: total z`)

(note: this is in effect what `sepov` does. Also see, e.g., `svylorenz`.)

# Estimation in Stata

Option 2 – sit on giant's shoulders: allow 'svy' prefix

Option 2: wrap the estimation step in a program and allow your program to work with the svy prefix

- ▶ compute  $\hat{\theta}$
- ▶ create a `predict ...`, `score` command which computes the  $\hat{Z}$
- ▶ then let svy do its magic or compute the variance 'manually' if user has no survey design feature (it's easy)

Note: Unfortunately, Stata (to date) does not permit programmer to know if user has used the svy prefix... that's a moderate annoyance!

# Example: An update to poverty

```
poverty varname [if] [in] [weight] [ ,  
    fracmedian(#) fracmean(#) line(#[varname]) ]
```

```
. svy : newpoverty nivie , fracmedian(.6)  
(running newpoverty on estimation sample)
```

Survey data analysis

```
Number of strata   =      21  
Number of PSUs    =      242  
Number of obs     =     5509  
Population size   = 15833.347  
Design df        =      221  
F( 0, 221)       =      .  
Prob > F         =      .
```

	nivie	Coef.	Linearized Std. Err.	t	P> t	[95% Conf. Interval]	
<b>fgt0</b>							
	_cons	.1928668	.0074118	26.02	0.000	.1782599	.2074737
<b>fgt1</b>							
	_cons	.0622773	.0042761	14.56	0.000	.0538503	.0707044
<b>fgt2</b>							
	_cons	.0310448	.0029002	10.70	0.000	.0253293	.0367604
<b>watts</b>							
	_cons	.0962296	.0079383	12.12	0.000	.0805851	.1118741
<b>pline</b>							
	_cons	12081.6	262.4697	46.03	0.000	11564.34	12598.86

## Illustration with ECHP data

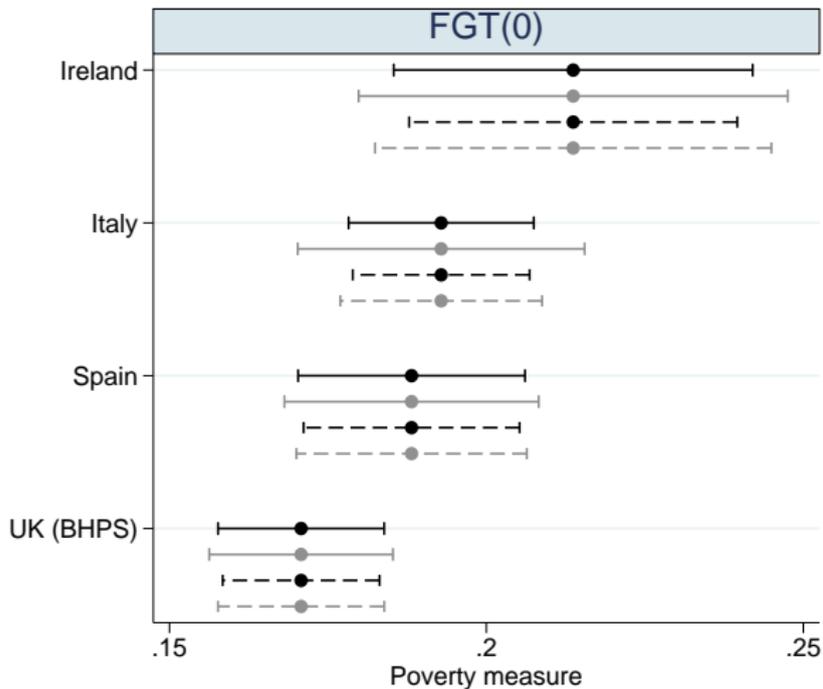
The European Community Household Panel survey is a (panel) survey

- ▶ 15 EU countries, 1994–2001
- ▶ stratified design (but info not always available)
- ▶ often clustered design (but info not always available)
- ▶ sample weights provided (but no details on construction)

Here, use only countries for which strata and PSU information is available, and only last wave.

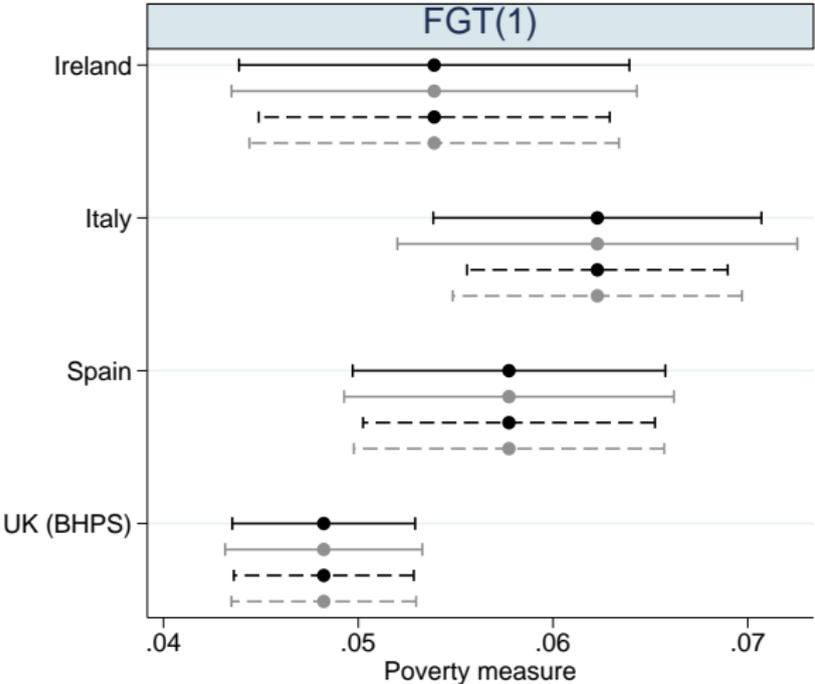
# Estimates of low income proportion

Poverty line estimated at 60% of median



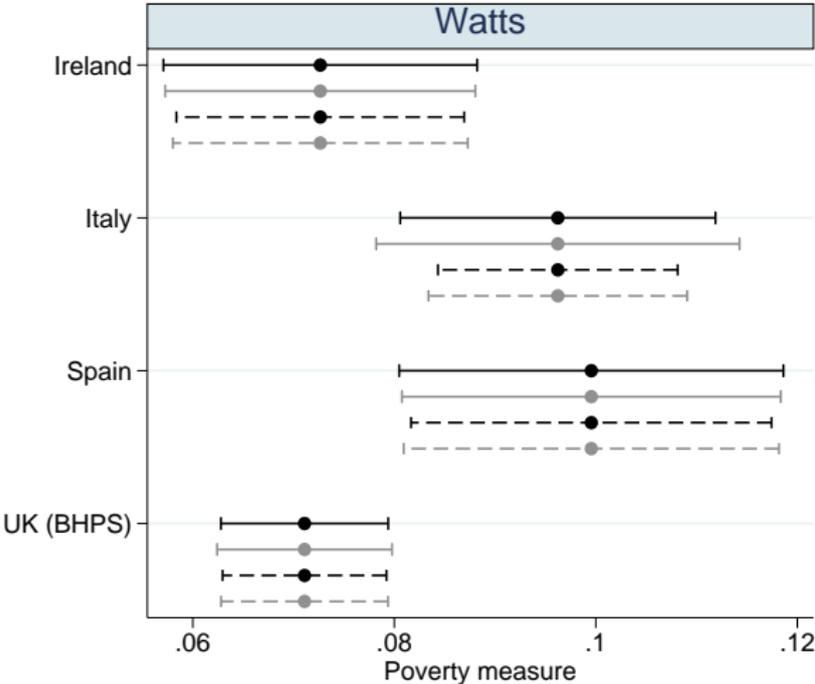
# Estimates of average income shortfall

Poverty line estimated at 60% of median



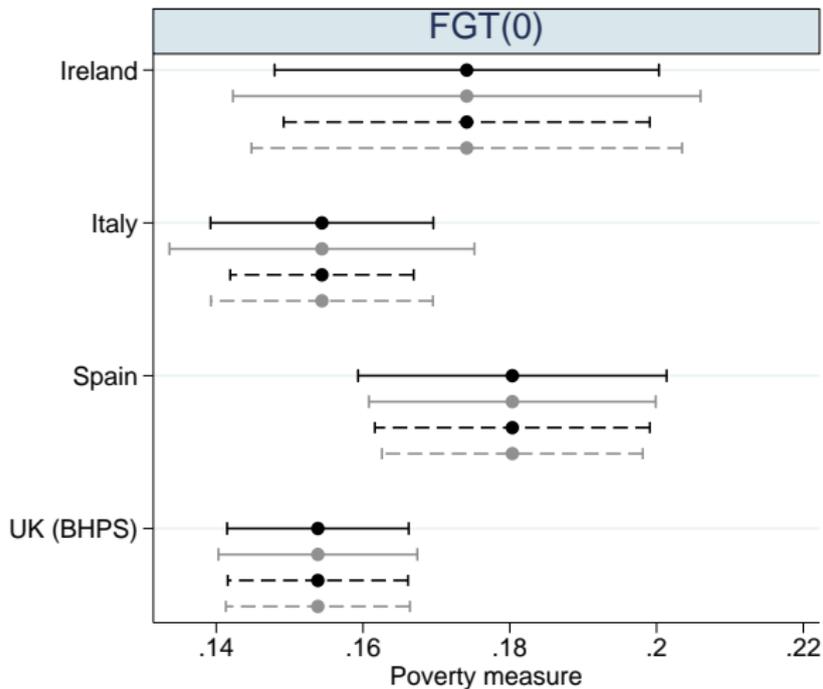
# Estimates of Watts index

Poverty line estimated at 60% of median



# Estimates of low income proportion (again)

Poverty line estimated at 50% of *mean*



## Summary and conclusion

Issue of estimation of poverty line is relevant to inference ...  
... but linearization methods available (IF approach is convenient and flexible)

Stata implementation in a command with `svy` prefix is not (very) difficult but pays off (all post-estimation commands become available, all configurations of `svyset` dealt with automatically)

Prototype of update to `poverty` almost ready.

# Acknowledgements

*This research is part of the MeDIM project (Advances in the Measurement of Discrimination, Inequality and Mobility) supported by the Luxembourg 'Fonds National de la Recherche' (contract FNR/06/15/08) and by core funding for CEPS/INSTEAD by the Ministry of Culture, Higher Education and Research of Luxembourg.*

*See <http://medim.ceps.lu> for more information.*