

 Istat

Integrating the Editing and Imputation of Incomes with the Record Linkage of Survey and Administrative Data: the Italian EU-SILC

Paolo Consolini and Silvano Vitaletti

*Eurostat and Statistics Finland
"International Workshop on Comparative EU Statistics on Income and Living Conditions: Issues and Challenges"*


Session 2: Editing, imputation (including imputed rent), gross/net conversion

Helsinki, 8th November 2006

Integrating the editing and imputation of incomes with record linkage...

DEALING WITH ADMINISTRATIVE SOURCE


- **Self-employment** (since 2004 survey)
- **Pensions** (since 2004 survey)
- **Employment** (since 2005 survey, in progress)
- **Other incomes** (in the future)



Integrating the editing and imputation of incomes with record linkage...

SELF-EMPLOYMENT INCOMES survey data collection


- **'Money drawn out' concept**
Earnings are 'money drawn out' from business for personal uses (consumption and saving), losses are 'money put into' it
- **Low degree of reticence by the interviewees**
The answers to the survey questionnaire do not involve tax consequences
- **The interviewer do not force the interviewee to respond when he seems embarrassed**
A non-response is better than a false response



Integrating the editing and imputation of incomes with record linkage...

SELF-EMPLOYMENT INCOMES survey data treatment

- **Detection and removal of unreliable observation based on interviewer's opinion**
Interviewers can observe directly the interviewees status and their behaviour towards the questionnaire
- **Detection and removal of outliers**
Hidiroglou-Berthelot method applied on univariate distribution of incomes (since 2005, incomes ratio between current year and the previous)
- **Missing (and unreliable) values imputation based on the huge stock of information collected on individuals and their households**
By means of IMPUTE module of IVEware, which uses a multivariate sequential regression approach to imputing item missing values



Integrating the editing and imputation of incomes with record linkage...

SELF-EMPLOYMENT INCOMES main results of treatment (2004 EUSILC)

Valid, missing and not reliable self-employment incomes in the survey data by professional status (a)

	Valid (%)	Missing (%)	Not reliable (%)	Outliers (%)	Total
No more self-employed	65.1	27.0	7.9	-	762
Co.Co.Co.	50.6	44.2	5.1	0.1	976
Entrepreneurs	80.3	6.4	13.3	-	593
Professionals	79.8	10.2	10.0	-	1,252
Artisans, shopkeepers, etc.	76.2	9.7	14.0	0.2	3,272
Co-helpers	54.5	36.6	8.9	-	191
Coop. stockholders	49.7	40.3	10.0	-	591
TOTAL	70.1	18.7	11.1	0.1	7,637

(a) At the moment of interview



Integrating the editing and imputation of incomes with record linkage...

SELF-EMPLOYMENT INCOMES main results of treatment (2004 EUSILC)

Mean self-employment incomes in the survey dataset by professional status (a)

	Valid	Imputed because missing	Imputed because not reliable or outlier		Total	
			Before	After	Before	After
No more self-employed	7,671	5,868	1,852	12,235	7,043	7,543
Co.Co.Co.	10,250	9,870	4,606	9,095	9,722	10,022
Entrepreneurs	22,622	22,231	7,309	18,190	20,442	22,006
Professionals	21,146	21,370	7,696	20,971	19,650	21,151
Artisans, shopkeepers, etc.	13,195	12,695	5,075	13,146	11,925	13,140
Co-helpers	13,436	12,161	6,176	10,025	12,416	12,665
Coop. stockholders	10,127	10,107	5,202	12,741	9,303	10,380
TOTAL	14,569	11,431	5,442	14,363	13,315	13,959


(a) At the moment of interview



Integrating the editing and imputation of incomes with record linkage...

Combining administrative and survey data

The term “record linkage” has been used to indicate the bringing together of two or more separately recorded pieces of information concerning a particular individual or family (H.B. Newcombe et al., 1985)



Integrating the editing and imputation of incomes with record linkage...

Combining administrative and survey data

Why do we combine administrative and survey data?


The aim is to improve the data quality on the income components (target variables) and on their earners by means of the imputation of the item non-response and the reduction of the measurement errors.

How is it possible to combine administrative and survey microdata?

By selecting an individual “matching-key” to link the same statistical unit among different data-sources.

Which is the best “matching-key” available?

The “tax number” (i.e. the personal identification number assigned to each person by the Italian Tax Authority)



Integrating the editing and imputation of incomes with record linkage...

Combining administrative and survey data

The process of the integration of micro-data from different sources consists of the following steps (P. van der Laan, 2000):

- harmonization of units:** are the statistical units defined uniformly in all sources? (comparability in space and time);
- harmonisation of reference periods:** do all data refer to the same period or the same point in time?
- completion of populations (coverage):** do all sources cover the same target population?
- harmonisation of variables/classifications:** are corresponding variables defined (classified) in the same way? (comparability in space and time);

Istat

Integrating the editing and imputation of incomes with record linkage...

Combining administrative and survey data

The integration of micro-data from different sources consists of the following steps (*continued*):

- adjusting for measurement errors (accuracy):** after harmonising definitions, do the corresponding variables have the same value?
- adjusting for missing data (item non-response):** do all the variables possess a value?
- derivation of variables:** are all variables derived using the combined information from different sources?
- checking overall consistency:** do the data meet the requirements imposed by identity relations?

Istat

Integrating the editing and imputation of incomes with record linkage...

Combining administrative and survey data

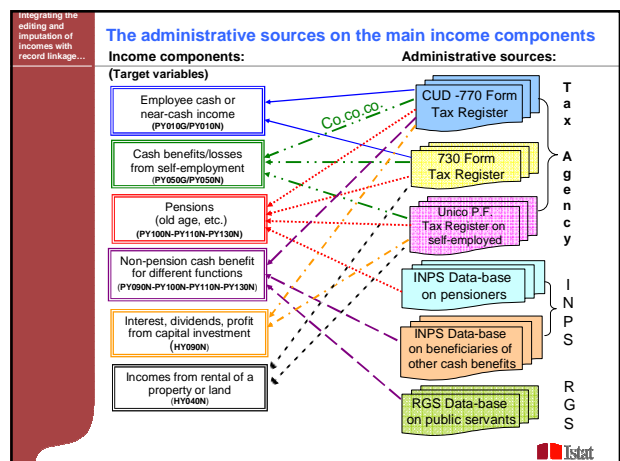
• MAIN RESULTS

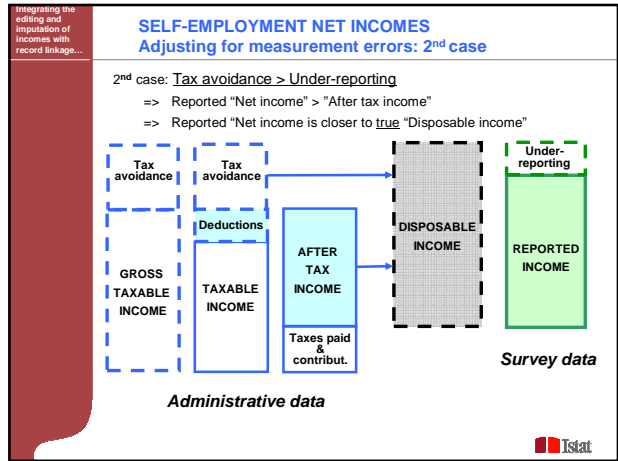
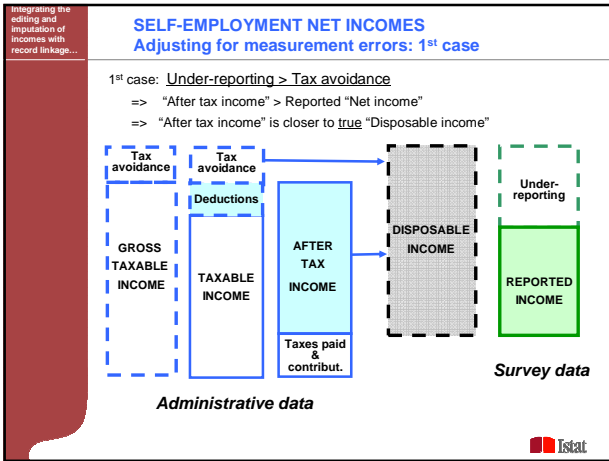
Coverage rate of the administrative source

	SURVEY DATA			
	Sampled	%	Interviewed	%
Linked with tax records	64,175	94.7	49,202	93.7
Not linked	3,568	5.3	3,307	6.3
TOTAL	67,743	100	52,509	100

The 93.7 % of the people interviewed has been successfully matched with Personal Tax Annual Register (embracing the universe of the Italian tax numbers).

Istat





Integrating the editing and imputation of incomes with record linkage...

SELF-EMPLOYMENT NET INCOMES

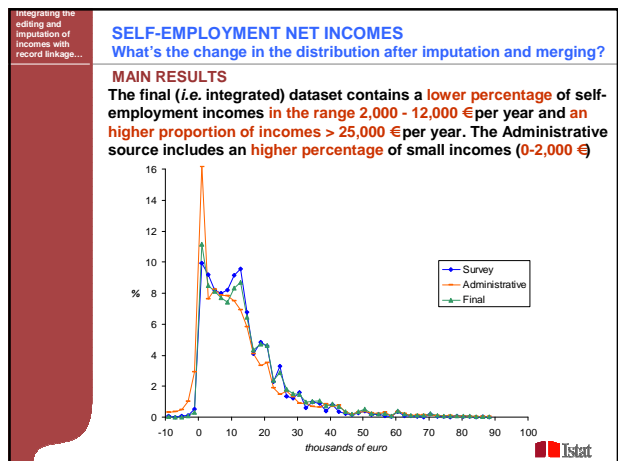
Adjusting for measurement errors

$$Y_{s-e} = \max(\text{Reported net income}, \text{After tax income})$$

Ho: both data-sources underestimate the true self-employment income

=> the measurement error is "minimised" by using the maximum value between the two data-sources.

Istat



SELF-EMPLOYMENT NET INCOMES

What's the change of the mean after imputation and merging?

MAIN RESULTS

- 1) The live-ware imputation has a **greater impact on:**
 - the mean income of **Cooperative stockholders (+12%)**
 - the mean income of **Artisans & shopkeepers (+10%)**
 - the mean income of **No more self-employed persons (+22%)**
- 2) Merging administrative and survey data has a **greater impact on:**
 - the mean income of **Entrepreneurs (+23%)**
 - the mean income of **No more self-employed persons (+22%)**
- 3) In the sample 1,00 persons, at present no more self-employed, **miss self-employment incomes (earned in the previous year).**

Professional status:	SURVEY DATA: before the imputation		SURVEY DATA: after the imputation		FINAL DATA: after the integration	
	Nr. Obs.	Mean of incomes	Nr. Obs.	Mean of incomes	Nr. Obs.	Mean of incomes
Not more self-employed	556	7,043	762	7,543	1,952	9,191
Co.Co.Co.	564	9,722	976	10,022	979	10,665
Entrepreneurs	535	20,442	593	22,006	593	27,071
Professionals	1,124	19,650	1,252	21,151	1,252	25,722
Artisans, shopkeepers etc.	2,950	11,925	3,272	13,140	3,274	15,740
Co-helpers	121	12,416	191	12,665	191	13,809
Coop. stockholders	353	9,303	591	10,380	591	12,011
Total	6,203	13,315	7,637	13,959	8,832	15,615

PENSIONS

Perceptients of pensions in the year 2003

DATA SOURCE	Perceptients of pensions (in thousands)				TOTAL
	Old-age & seniority	Survivors	Inability & Disability ^(a)	Social	
Eu-Silc 2004 - Survey data	10.672	3.375	2.210	1.021	14.668
Eu-Silc 2004 - Merged data	10.573	4.194	4.148	749	15.861
Social Security Database ^(b)	10.471	4.602	4.523	743	15.726

(a) except for the assistants' allowance
 (b) to compare with the EU-SILC, only adults perceptients have been included

SELF-EMPLOYMENT INCOMES

MAIN RESULTS

	survey	tax	all
Employees	36.4	63.6	100.0
Entrepreneurs	79.3	20.7	100.0
Professionals	71.2	28.8	100.0
Artisans/shopkeepers...	73.9	26.1	100.0
Co-helpers	84.6	15.4	100.0
Coop. stockholders	88.0	12.0	100.0
Co.co.co.	94.3	5.7	100.0
Unemployed	44.7	55.3	100.0
Other inactive	28.3	71.7	100.0
All	68.2	31.8	100.0

In the final data, the majority of retained records for these sub-groups ('employees', 'unemployed' and 'other inactive') comes from the tax datasource