The Italian n**ew** survey on Enterprises Final Consumption of Energy Products (COEN) – 2011: an innovative editing procedure

Giovanni Seri, Roberta Varriale, Ugo Guarnera
Istat

Nuremberg, 11/09/2013

Istat

Introduction

Regulation (EC) No 1099/2008, successively amended by the Commission Regulation (EU) n.844/2010 (Regulation in the following), establishes a common framework for the production, transmission, evaluation and dissemination of comparable energy statistics in the Union

- Ministry of economic development (**MISE**)

- Italian National Agency for New Technologies, Energy and Sustainable Economic Development (**ENEA**)

 - The Italian National Statistical Institute (**Istat**): **survey on Enterprises Final Consumption of Energy Products (COEN)**

- *main characteristics of the COEN survey*

- *innovative procedure of data checking/editing*

- *conclusions*

- ISPRA Institute for Environmental Protection and Research (Emission Trading System)

- TERNA (Electricity Infrstructure)

- GSE Energy Services (supports renewable energy sources)

Istat

The survey on Enterprises Final Consumption of Energy Products (COEN)

COEN survey aims at estimating the enterprises consumption of a list of products belonging to the following categories:

-       Electricity and heat
-       Natural gas
-       Oil and petroleum products
-       Solid fossil fuels and manufactured gases
-       Renewable energy and energy from waste.

Reference year 2011

Web data collection

Note: energy statistics in PRODCOM until 2010

Istat

The survey on Enterprises Final Consumption of Energy Products (COEN)

Final energy consumption
(end-use sectors):

Iron and Steel
Chemical (including Petrochemical)
Non-Ferrous Metals
Non-Metallic Minerals
Transport Equipment
Machinery
Mining and Quarrying
Food, Beverages and Tobacco
Pulp, Paper and Printing
Wood and Wood Products (other than pulp and paper)
Construction
Textile and Leather
Not Elsewhere Specified – Industry

Rail
Domestic Navigation
Road
Pipeline Transport
Aviation

Commercial and Public Service

Istat

The survey on Enterprises Final Consumption of Energy Products (COEN)

Sampling design: one stage stratified simple random sample
Population: enterprises with at least 2.5 employees (ASIA archive)

NACE sector	Population	Sample	Response rate
Industry	367K	17.2k	40.7%
Transport	30k	5.1k	30.9%
Commercial and Public Service	700k	18.3k	36.8%
Total	1097k	40.7k	37.1%

Istat

The data editing procedure

Data editing strategy:

- mixture modeling to detect unity measure errors → Electricity, Gas, Diesel oil and Heat

- selective editing of influential errors → Electricity and Gas

- deterministic procedures to remove errors → all products

Software:
mixture modeling → R program
selective editing → R package Selemix
deterministic procedures → SAS program

13k observations collected by October 2012 and used to produce preliminary estimates

The data editing procedure: Electricity

The three main sector, Industry, Transport and Services, treated separately

Sector/Response pattern	Respondents (10 Oct 2012)	Zero consumption	Quantity only	Expenses only	Both Quantity and Expenses	PRODCOM 2010
Industry	6117	75	157	80	5805	**4406**
Transport	1243	264	70	28	881	
Services	5608	160	245	201	5002	
Total	12968	499	472	309	**11688**	

11688 respondents out of 12968 filled in both Quantity and Expenses

In the Industry sector, 4406 out of 5805 enterprises also responded to the PRODCOM survey in 2010 on Electricity consumption (Quantity and Expenses)

Note: Electricity consumption was a mandatory data

Istat

Mixture modeling to detect unity measure errors: Industry sector - Electricity

5805 units respondent to both Quantity and Expenses. The procedure has been applied separately to the units with auxiliary information from the PRODCOM survey 2010 (4406) and to the units without auxiliary information (1399)

Quantity could be indicated in: kilowatt (kW), megawatt (MW) gigawatt (GW) or Gigajoul (Gj)

Expenses were asked in thousands of euro

(after a preliminary analysis) 5 unity measure error pattern:

>   1. no errors in both Quantity and Expenses
>
>   2. Expenses expressed in euros instead of thousands of euros
>
>   3. Quantity expressed in MW or GW instead of kW and MW, respectively
>
>   4. Quantity expressed in kW or MW instead of MW and GW, respectively
>
>   5. a combination of error types 2 and 4

Each unit has been classified according to its error pattern (and errors have been edited) using a classification model-based approach (mixture models)

Istat

Mixture modeling to detect unity measure errors

Clusters of units by consumption of Electricity: Quantity and Expenses. Industry sector, log scale.
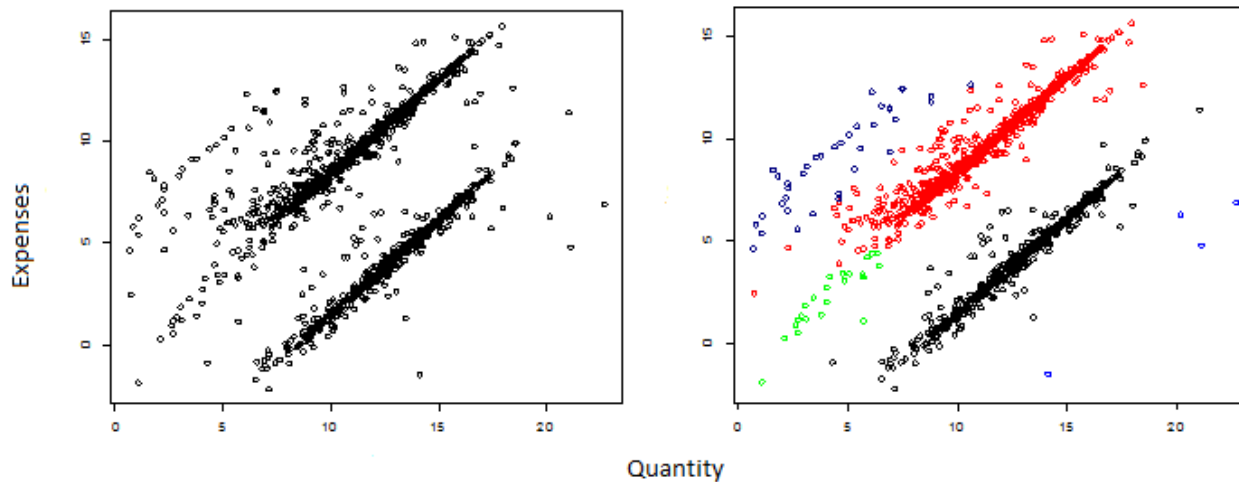
Cluster	Nr of units	No PRODCOM	PRODCOM
1	3343	572	2771
2	2092	763	1329
3	32	4	28
4	247	23	224
5	91	37	54
Total	**5805**	**1399**	**4406**

Selective editing of influential errors based on contamination models

Units also respondents to PRODCOM survey were treated with **Selemix**: information stemming from PRODCOM have been used as auxiliary variables in the contamination model

Response variables:
    Quantity, Expenses

Auxiliary variables:
    Quantity in PRODCOM 2010, Expenses in PRODCOM 2010, Class size

Results:
    In the Industry sector 80 observations (out of 4406) marked as 'influential' and corrected.

Istat

Deterministic editing

A set of 11239 'good observations' was identified (complete responses, 'reasonable' unit price)

Intervals of 'acceptance' of Electricity consumption were defined taking into account combinations of NACE domains and size classes.

Units non respondent to PRODCOM 2010 but providing at least Quantity to COEN (1399+157) were checked: those with consumption external to the interval were clerically checked (<50 observations were corrected).

Units providing Expenses only (80) were first imputed using the median unit price per combination of NACE domains and size classes and then checked deterministically using the interval from the set of 'good observations'.

Istat

Deterministic editing

For all the other products (except Electricity, Gas, Diesel oil and Heat), the editing process was fully deterministic based on the following steps:

(a)   unit price belongs to a given interval predefined;

(b)   unit price belongs to a given interval predefined after expenses are divided by 1000;

(c)   enterprises with too low consumption were marked and clerically checked;

(d)   respondent to Expenses only were imputed using the median unit price in a set of 'similar' observations.

Istat

Deterministic editing

More checking

- main contributors to final estimates were clerically checked

- estimates were compared to value stemming from external sources (International Energy Agency 2010, Terna 2010-2011)

- estimates were discussed with experts in the field

Istat

Conclusions

COEN survey can be considered an experimental survey on energy statistics

Through this experience, Istat is planning to conduct a steady survey in the future

The results can be considered quite good with respect to collaboration with enterprises, response rate and evaluation of the subject matter experts

The analysis of the raw and treated data and the editing procedure highlighted some features of the survey that need to be revised in the future (for example unity measure)

Individual comparisons with units subject to ETS to integrate COEN data
Analysis of pattern of response to reorganize questionnaire and sampling strategy

Istat

Further analysis:

- Suggestions?

Thank you for your attention
seri@istat, varriale@istat.it, guarnera@istat.it

Istat