

Non sampling errors in the consumer prices survey

Errori non campionari nell'indagine sui prezzi al consumo

Matteo Mazziotta

Alessandro Pallara

Adriano Pareto

Angela Claudia Seeber

Istat, via Balbo, Roma, mazziott@istat.it

Riassunto: Questo lavoro si propone di studiare la relazione tra qualità dei dati e tecniche d'indagine adottate nell'ambito delle rilevazioni territoriali dell'indagine Istat sui prezzi al consumo. A tale fine, sono stati analizzati congiuntamente i risultati di un'analisi sui microdati dell'indagine sui prezzi al consumo e i dati provenienti da una rilevazione presso gli Uffici Comunali di Statistica, relativa alle tecniche d'indagine utilizzate per effettuare la rilevazione.

Keywords: consumer price index, data quality, data collection techniques, error risk.

1. Introduction

The Consumer Price Index (CPI) is the main measure of price inflation in Italy. During the last few years the Istituto Nazionale di Statistica (Istat) has undertaken a number of developments to improve the quality of CPI production process. The quality management program includes a re-design of the sampling selection for locally collected prices, exploitation of hand-held computers for local data collection, a re-engineering of computerized systems for data processing, a re-development of central price collection and related index production processes. As a part of these endeavors to strengthen data quality, a survey of local price collection offices has been carried out in 2005 in order to yield information related to survey techniques adopted for local data collection during 2004. The results of that survey are analyzed in this paper jointly with locally collected prices during 2004, aiming at providing some insight on the relation between data quality and data collection techniques. As for non sampling errors, we analysed a substantial part of microdata collected in 2004, containing elementary price quotes as well as supplementary information on imputed prices, reduced prices and substitutions of outlets or references. On the other hand, local data collection methods were investigated through the following items, explored through the responses to survey of local price collection offices: field workers selection criteria (FS), feedbacks between local office managers and field workers (FC), field workers training (FT), use of auxiliary forms for data collection (AF), course of action in case of errors by field workers (CA), field workers rotation on outlets (RF), outlet selection methods (OS), field worker autonomy in substituting (items/outlets) (AS), mode of first contact with outlets (FO), advance letter to outlets (LO), use of pocket computers for data capturing (PC), collection methods of fees for some services (SF), outlet sample controls (OC), data checks (CD), use of own supplementary software for data entry (SW).

2. Methods and results

Analysing monthly and annual frequencies of repeated prices and reported special events (imputed prices for non-observed outlets/references, reduced prices, substitutions), three distinct groups were defined: a group of 19 cities where different outlying frequencies were observed (OL), a “standard” group of 62 cities characterized by very low frequencies of irregular prices (STD) and a small group of seven cities registering events of all the above types, therefore representing a possible benchmark (BM) for error risk management. In order to strengthen this interpretation, quality groups have been related to the variables describing the survey technique, obtained after re-arranging of original responses to survey of local price collection offices. Table 1 shows the relative cross tabulation and a synthetic representation of the main results of a multiple correspondence analysis which uses data quality group classification as illustrative variable.

Table 1: *Relation between quality groups and survey technique variables*

-----					-----				
Item	Modalities: position on first two axes (**)	OL - -, +	STD 0, 0	BM ++, -	Item	Modalities: position on first two axes (**)	OL - -, +	STD 0, 0	BM ++, -
FS	Exams or titles: ++, -	3	15	3	FO	Announce in advance: +, -	6	14	5
	No exams or titles: -, -	11	46	4		Visit without announce: - -, -	0	5	0
	Field work contracted out: - -, ++	1	1	0		Directly data collection: -, +	9	16	2
FC (*)	Direct with form checks: +, -	8	49	7	LO	Istat and local letter: ++, +	2	19	3
	Only form checks: - -, -	2	2	0		Only Istat letter: +, - -	3	5	1
	No checks or no contact: -, ++	5	11	0		Only local letter: -, +	9	31	3
FT	By Istat and local: ++, +	2	15	3		No letter at all: -, -	1	7	0
	Only by Istat: -, -	8	32	1	PC	All by pocket computers: -, +	0	5	1
	Only local: +, +	0	8	2		In part by pocket computers: ++, ++	1	13	1
	No training: - -, +	5	7	1		Only paper forms: -, -	14	44	5
AF	Own auxiliary forms: +, +	6	28	5	FS (\$)	Mainly by published price list: +, -	10	47	6
	No auxiliary forms:	9	34	2		Other or non response: - -, +	5	15	1
CA	Training on field: ++, -	0	6	1	OC	Direct and by telephone: +, - -	3	20	3
	Extra theoretical training: +, -	4	20	3		Only by telephone: +, +	3	4	3
	Control on field: +, ++	2	7	0		Only direct: -, ++	4	15	0
	Other or no measures: -, -	9	29	3		No outlet control: - -, -	5	23	1
RF	Rotation of field workers: - -, ++	4	11	1	DC	Documented data checks: ++, +	1	22	3
	No rotation of field workers: +, -	11	51	6		No documented data checks: +, -	7	29	4
OS	Outlet frames & statistics: ++, ++	1	6	0		No data checks at all: - -, +	7	11	0
	Only outlet frames: ++, -	4	16	3	SW	Own extra software: ++, ++	1	6	2
	Only statistics: - -, ++	1	8	1		No extra software: -, -	14	56	5
	Field workers without data: -, -	5	20	2					
	Expert use or non response: -, +	5	12	0					
AS (\$)	O-office, R-field worker: ++, -	0	10	4					
	All by local head office: +, +	6	11	0					
	All by field workers: -, +	7	38	3					
	Other or non response: -, -	2	3	0					

(*) for example control of outlet reference person's signature

(\$) O = outlet, R = reference

(#) excluding paper forms for housing and service fares

(\$) ≥ 7 of 9 services by direct, internet or legislation checks

(**) +: 0-0,5 / ++: >0,5 / -: -0,5-0 / --: <-0,5

The results shows that data quality is positively correlated with the first factor, which measures the quantity, accuracy and differentiation of data collection features. The second factor may rather be interpreted as a measure of modernity in instruments adopted for data collection and carrying of survey operations.

On the whole, there seems to be some degree of association between differences in data quality and the heterogeneity in techniques used for data collection, suggesting that the generally observed negative performances were mainly due to organizational problems of field work, which not necessarily might be resolved by adopting modern instruments for data collection and definitely not by breaking apart local office managers and field workers.