Istat
Istituto Nazionale
di Statistica

Community Innovation Survey CIS2008

2006-2008

File description

December 2010

CONTENTS

1. Introduction

In order to satisfy the users' needs, Istat disseminates microdata files for scientific Research (MFR). The dissemination of the statistical information must be performed in full compliance with the regulations pertaining the privacy of respondents. For the Community Innovation Survey CIS3 and CIS4, a statistical disclosure control methodology was set-up and two MFR were released on the Istat web-site. For CIS2008, the same methodology was applied.

In section 2 useful information about the status of each variable in the MFR is provided. The statistical disclosure control methodology is briefly described in Section 3. In Section 4, the main effects of the applied statistical disclosure limitation methodology are illustrated. These may be used to assess the impact of the applied protection methods.

In this document only the modifications due to the statistical disclosure limitation methodology are described.

The users of the microdata file for scientific research must comply with the conditions of the signed agreement.

2. Description of variables

The protection of the confidentiality of respondents was achieved by modifying the information content of some variables. This section summarizes only the changes due to the statistical disclosure limitation methodology.

The MFR may be disseminated as a tab-delimited file or as a fixed length format.
The first row of the tab-delimited file contains the names of the variables.
The decimal separator is ".".
The character "." means "missing data or not applicable".
The MFR contains information on enterprises.
Each record describes an enterprise.
The MFR contains 19904 records and 141 variables.
The sum of the weights (variable *weightnr*) equals 208637.531.

Table 1. The status[1] of each variable in the Italian CIS2008 MFR.

Variable	Question	SDC status	Short description
id	Name of the enterprise	changed	removed
NUTS	Address	changed	aggregated at national level
nace	Main economic activity	changed	NACE.rev2 2-digit, NACE05-09 = NACE "B"; NACE11-12 = NACE "11-12"; NACE38-39 = NACE "38-39"
HO	Country of head office	changed	".": no answer; 1: Home country; 2: Europe+Candidate Countries + EFTA countries; 3: rest of the world; NN: the enterprise does not belong to a group;
rrdinx	Expenditure in intramural R&D	changed	0.09% of records changed
rrdexx	Expenditure in extramural R&D	changed	0.07% of records changed
rmacx	Expenditure in acquisition of machinery	changed	0.08% of records changed
roekx	Expenditure in other external knowledge	changed	0.06% of records changed
rprex	Expenditure in other preparation	changed	0.07% of records changed
rtrx	Expenditure in training	changed	0.09% of records changed
rmarx	Expenditure in market introduction of innovation	changed	0.07% of records changed
rtot	Total for four expenditure categories	changed	0.11% of records changed
turn06	Total turnover in 2006	changed	24.43% of records changed
turn08	Total turnover in 2008	changed	24.30% of records changed
EMP06	Total number of employees in 2006	changed	**"10-49":** 0-49 employees; **"50-249":** 50-249 employees; **"250+":** more than 250 employees; **"50+":** only for NACE 45, 65, 66, 68, 72, 77 and B, category "50-249"and category 250+" are aggregated into a new one; **"ALL":** only for NACE 53 all categories are aggregated into a single one.

---

[1] The status refers **only** to the application of the statistical disclosure control (SDC) methodology.

			"10-49": 0-49 employees; **"50-249":** 50-249 employees; **"250+":** more than 250 employees; **"50+":** only for NACE 45, 65, 66, 68, 72, 77 and B, category "50-249"and category 250+" are aggregated into a new one; **"ALL":** only for NACE 53 all categories are aggregated into a single
EMP08	Total number of employees in 2008	**changed**	one.
STRB	Stratum of the enterprise when sampling	**changed**	Removed
STRA	Stratum of the enterprise according to the questionnaire	**changed**	Removed
weight	Weighting factors	**changed**	Removed
weightnr	Weights corrected for non-response bias	not changed	

In the next section, more details on the information content of the changed variables are given.

3. Statistical disclosure control methodology

3.1 Disclosure Scenarios
The units at risk of identification were singled out by means of two disclosure scenarios: one based on structural information and the other one derived from non-structural information. The adopted criteria were a mixture of frequency rules, density-based clustering algorithms and dominance criteria.

3.2. Changes on Variables

3.2.1. Variable Suppression
The following variables were removed from the microdata file:

    a. Name (variable *id*)

    b. Stratum Information (variables *STRA* and *STRB*)

    c. Base weight (variable *weight*). Observation: the final calibrated weight (variable w*eightnr*) was not changed.

When a variable was removed, its values were replaced by the missing value symbol, i.e. ".".

3.2.2. Global Recoding

Global recoding is one of the simplest and well-known statistical disclosure limitation methodologies. It can be applied only to categorical variables. Recoding consists in merging categories of the categorical variable. Obviously, there is a difference between the old and new categories.

For the release of the Italian CIS2008 microdata file for scientific research, the following variables were recoded:

1. *Principal economic activity* (*nace*) was recoded as follows:
   a. NACE was aggregated in NACE at 2 digits, according to NACE rev2;
   b. NACE 05, 06, 07, 08 and 09 were aggregated into a single category called "NACE B";
   c. NACE 11 and 12 were aggregated into a single category called "NACE 11-12"
   d. NACE 38 and 39 were aggregated into a single category called "NACE 38-39"

2. *Number of employees* (*EMP08*) was recoded in three main categories. For the Italian CIS2008 microdata file for scientific research, the following categories were then obtained:
   a. small size: 10 – 49 employees, category "10-49"
   b. medium size: 50 – 249 employees, category "50-249"
   c. large size: more than 250 employees, category "250+"
   d. only for NACE 45, 65, 66, 68, 72, 77 and B, category "50-249"and category 250+" are aggregated into a new one "50+";
   e. only for NACE 53 all categories of *EMP08* were aggregated into a single one, category "ALL"

3. *Number of employees at the first year of the reference period* (*EMP06*) was recoded in three main categories. For the Italian CIS2008 microdata file for scientific research, the following categories were then obtained:
   a. small size: 10 – 49 employees, category "10-49"
   b. medium size: 50 – 249 employees, category "50-249"
   c. large size: more than 250 employees, category "250+"

d. only for NACE 45, 65, 66, 68, 72, 77 and B, category "50-249"and category 250+" are aggregated into a new one "50+";

e. only for NACE 53 all categories of *EMP08* were aggregated into a single one, category "ALL"

4. *NUTS* categories were aggregated into a single one; the geographical location of enterprise is released only at national level.

5. *HO* categories were aggregated into the following categories: 1 (Home), 2 (Europe + Candidate Countries + EFTA countries), 3 (rest of the world), NN (the enterprise does not belong to a group), and 0 (no answer).

3.2.3. Significant Digits in Continuous Variables

The numerical variables were rounded to unity. Two decimals were preserved only for variables *TURNMAR, TURNIN* e *TURNUNG*. Three decimals were preserved only for the weights (variable *weightnr*).

3.3. Protection of Enterprises

A further protection level was achieved by perturbing the numerical variables considered identifying in either disclosure scenario. Only the records of the units at risk of re-identification were modified and only for the key or sensitive variables. These modifications concern:

a) the *turn08* values of the units at risk in the scenario based on structural information

- For each combination of *nace* and *EMP08*, the *turn08* values of the units at risk were modified by imputing the values from the nearest safe unit. For each combination of *nace* and *EMP08*, on the tails of the *turn08* distribution an individual ranking was applied. The perturbation applied to *turn06* was generally proportional to the perturbation of *turn08*.

b) the values of *rrdinx, rrdexx, rmacx, roekx, rprex, rtrx, rmarx, rtot* of the units considered at risk in the identification scenario based on non-structural information

- For the units at risk, *rtot* and its components were modified. Except for very few units, the perturbation was proportional to the perturbation of *turn08*. The same multiplicative perturbation was applied to all the components of *rtot*.

4. Information content analysis

When applying whatever statistical disclosure limitation methodology, some information loss is unavoidable. (otherwise it wouldn't be possible to guarantee the confidentiality of the respondents)

For the Italian CIS2008 microdata file for scientific research, all weighted totals of *Turn08* were exactly preserved, for all combinations of the categorical identifying variables (*Nace* and *Emp08*).

The weights (variable *weightnr*) were not modified.

4.1. Number of modified turn08 values

With respect to the spontaneous identification scenario based on structural information, the only perturbed variable was *turn08*. The perturbation was applied only to the units at risk. The safe units were generally released with their original values. Table 2 presents, for each combination of *nace* and *EMP08*, the percentage of modified *turn08* values.

Table 2. Percentage of modified units.

nace	EMPp08	Observations	Percentage of modified units	nace	EMP08	Observations	Percentage of modified units
10	250+	63	30.16	37	10-49	48	25.00
10	50-249	128	21.88	37	50-249	11	100.00
11-12	10-49	73	26.03	38-39	10-49	244	22.13
11-12	250+	11	100.00	38-39	250+	41	53.66
11-12	50-249	24	33.33	38-39	50-249	89	32.58
13	10-49	181	23.76	41	10-49	1546	30.47
13	250+	33	42.42	41	250+	21	33.33
13	50-249	91	23.08	41	50-249	160	28.75
14	10-49	282	10.28	42	10-49	242	29.75
14	250+	34	32.35	42	250+	22	40.91
14	50-249	81	29.63	42	50-249	83	33.73
15	10-49	137	25.55	43	10-49	2149	26.06
15	250+	10	100.00	43	250+	12	100.00
15	50-249	57	8.77	43	50-249	133	28.57
16	10-49	280	11.07	45	10-49	998	24.75
16	250+	9	100.00	45	50+	130	10.00
16	50-249	50	30.00	46	10-49	845	23.43
17	10-49	105	13.33	46	250+	89	20.22
17	250+	19	57.89	46	50-249	198	24.24
17	50-249	66	10.61	47	10-49	861	12.66
18	10-49	370	10.27	47	250+	152	21.05
18	250+	11	100.00	47	50-249	164	25.00

| 18 | 50-249 | 34 | 47.06 | 49 | 10-49 | 606 | 24.09 |
19	10-49	29	51.72	49	250+	93	39.78
19	250+	11	100.00	49	50-249	94	32.98
19	50-249	16	25.00	50	10-49	23	17.39
20	10-49	56	12.50	50	250+	22	27.27
20	250+	41	24.39	50	50-249	19	26.32
20	50-249	52	19.23	51	10-49	22	54.55
21	10-49	13	100.00	51	250+	6	100.00
21	250+	39	25.64	51	50-249	2	100.00
21	50-249	24	29.17	52	10-49	143	26.57
22	10-49	248	27.02	52	250+	86	10.47
22	250+	40	25.00	52	50-249	118	25.42
22	50-249	157	25.48	53	ALL	21	42.86
23	10-49	253	12.25	55	10-49	607	12.85
23	250+	51	23.53	55	250+	21	19.05
23	50-249	75	30.67	55	50-249	96	10.42
24	10-49	105	11.43	56	10-49	644	25.93
24	250+	42	16.67	56	250+	41	14.63
24	50-249	105	37.14	56	50-249	64	12.50
25	10-49	538	30.30	58	10-49	56	8.93
25	250+	51	45.10	58	250+	12	100.00
25	50-249	147	25.85	58	50-249	31	12.90
26	10-49	96	28.13	59	10-49	42	14.29
26	250+	28	42.86	59	250+	3	100.00
26	50-249	38	13.16	59	50-249	9	100.00
27	10-49	91	27.47	61	10-49	19	21.05
27	250+	46	10.87	61	250+	15	100.00
27	50-249	58	24.14	61	50-249	9	100.00
28	10-49	242	11.16	62	10-49	144	32.64
28	250+	121	21.49	62	250+	50	10.00
28	50-249	95	26.32	62	50-249	81	28.40
29	10-49	33	12.12	63	10-49	110	23.64
29	250+	59	30.51	63	250+	12	100.00
29	50-249	26	38.46	63	50-249	37	18.92
30	10-49	24	20.83	64	10-49	140	23.57
30	250+	24	12.50	64	250+	163	15.34
30	50-249	17	52.94	64	50-249	184	29.89
31	10-49	178	25.84	65	10-49	27	40.74
31	250+	19	31.58	65	50+	66	22.73
31	50-249	86	9.30	66	10-49	189	19.58
32	10-49	206	16.99	66	50+	34	11.76
32	250+	12	100.00	68	10-49	128	26.56
32	50-249	59	11.86	68	50+	24	58.33
33	10-49	351	23.36	71	10-49	186	10.22
33	250+	9	100.00	71	250+	15	100.00
33	50-249	47	14.89	71	50-249	54	12.96
35	10-49	106	42.45	72	10-49	45	37.78
35	250+	29	34.48	72	50+	20	55.00
35	50-249	49	36.73	77	10-49	72	11.11
36	10-49	25	36.00	77	50+	19	31.58
36	250+	16	25.00	B	10-49	172	31.98
36	50-249	39	20.51	B	50+	23	21.74

4.2. Variance Comparison

The ratios between the *turn08* variance after and before protection are shown in Table 3 and in Figure 1.

Table 3. Ratios of *turn08* variances.

nace	EMP08	Variaces ratio	nace	EMP08	Variances ratio
10	250+	0.94	37	10-49	1.06
10	50-249	0.99	37	50-249	0.70
11-12	10-49	0.96	38-39	10-49	1.01
11-12	250+	0.86	38-39	250+	0.84
11-12	50-249	1.12	38-39	50-249	0.87
13	10-49	0.83	41	10-49	1.03
13	250+	0.81	41	250+	0.88
13	50-249	1.12	41	50-249	0.92
14	10-49	0.55	42	10-49	0.92
14	250+	1.01	42	250+	0.88
14	50-249	1.01	42	50-249	0.99
15	10-49	1.12	43	10-49	0.93
15	250+	0.76	43	250+	0.47
15	50-249	0.98	43	50-249	0.97
16	10-49	0.93	45	10-49	1.00
16	250+	0.81	45	50+	0.77
16	50-249	0.83	46	10-49	1.01
17	10-49	1.00	46	250+	1.07
17	250+	0.47	46	50-249	0.97
17	50-249	0.88	47	10-49	0.88
18	10-49	1.00	47	250+	1.02
18	250+	0.61	47	50-249	0.94
18	50-249	1.33	49	10-49	0.99
19	10-49	0.80	49	250+	0.56
19	250+	0.78	49	50-249	1.23
19	50-249	0.50	50	10-49	0.43
20	10-49	1.04	50	250+	0.78
20	250+	0.58	50	50-249	0.44
20	50-249	0.37	51	10-49	1.03
21	10-49	1.26	51	250+	0.32
21	250+	1.07	51	50-249	0.00
21	50-249	1.03	52	10-49	1.02
22	10-49	1.01	52	250+	0.84
22	250+	0.68	52	50-249	0.92
22	50-249	1.00	53	ALL	1.03
23	10-49	1.06	55	10-49	0.79
23	250+	0.87	55	250+	0.86
23	50-249	0.97	55	50-249	0.84
24	10-49	1.01	56	10-49	0.95
24	250+	0.88	56	250+	0.78
24	50-249	0.98	56	50-249	1.38
25	10-49	0.62	58	10-49	0.91
25	250+	1.03	58	250+	0.71
25	50-249	0.97	58	50-249	0.94
26	10-49	0.85	59	10-49	0.73
26	250+	1.00	59	250+	0.00
26	50-249	0.54	59	50-249	0.72

27	10-49	0.79	61	10-49	0.34
27	250+	0.83	61	250+	0.67
27	50-249	0.99	61	50-249	0.89
28	10-49	0.60	62	10-49	0.81
28	250+	0.98	62	250+	0.81
28	50-249	1.07	62	50-249	0.77
29	10-49	0.31	63	10-49	1.00
29	250+	0.36	63	250+	0.87
29	50-249	0.66	63	50-249	1.05
30	10-49	0.68	64	10-49	0.99
30	250+	0.84	64	250+	0.79
30	50-249	0.87	64	50-249	0.97
31	10-49	0.94	65	10-49	1.10
31	250+	0.85	65	50+	0.95
31	50-249	0.92	66	10-49	0.99
32	10-49	0.90	66	50+	0.89
32	250+	0.80	68	10-49	0.88
32	50-249	0.94	68	50+	0.74
33	10-49	1.03	71	10-49	0.58
33	250+	0.48	71	250+	0.69
33	50-249	0.84	71	50-249	1.84
35	10-49	0.99	72	10-49	1.01
35	250+	1.05	72	50+	0.77
35	50-249	1.09	77	10-49	1.01
36	10-49	0.41	77	50+	1.00
36	250+	0.57	B	10-49	0.87
36	50-249	1.00	B	50+	0.10

Figure 1. Variance comparison of *turn08* before and after protection.
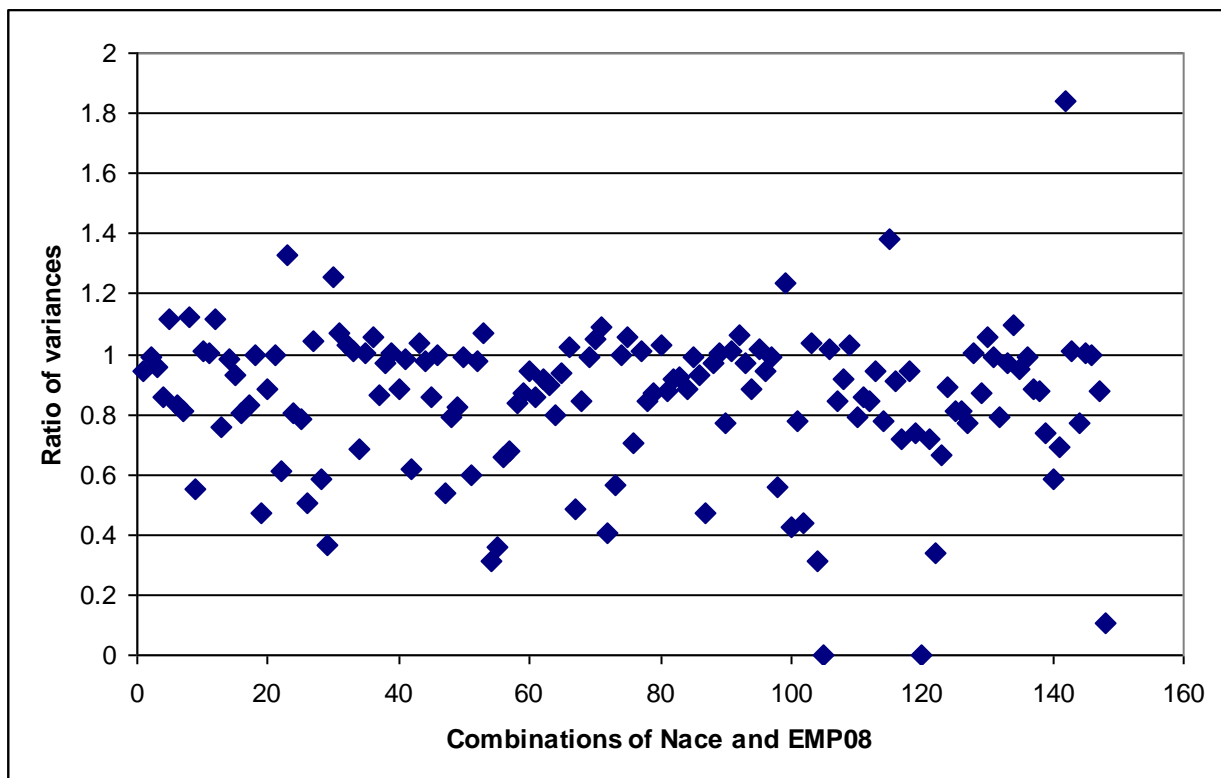
4.3. Correlations Comparison

In Table 4 the correlations between the original and perturbed *Turn08* values, between original *turn08* and *rtot* and between perturbed *turn08* and *rtot* are shown. A similar effect was observed on the other components of *rtot*. The missing values are due to the extremely reduced (1 or 2) number of non-missing values.

Table 4. Coefficients of correlations.

nace	EMP08	cor(*turn08,turn08Protected*)	cor(*ttot,ttotProtected*)	cor(*turn08,ttot*)	cor(*turn08Protected,rtotProtected*)	nace	EMP08	cor(*turn08,turn08Protected*)	cor(*rtot,rtotProtected*)	cor(*turn08,rtot*)	cor(*turn08Protected,rtotProtected*)
10	10-49	0.97	1	0.19	0.16	36	50-249	1	1	0.14	0.16
10	250+	1	1	0.69	0.69	37	10-49	0.98	1	0.25	0.24
10	50-249	1	1	0.3	0.3	37	50-249	0.81	1	0.3	0
11-12	10-49	1	1	0.44	0.43	38-39	10-49	1	1	0.11	0.11
11-12	250+	0.93	1	0.7	0.59	38-39	250+	0.91	1	0.24	0.26
11-12	50-249	0.99	1	0.39	0.37	38-39	50-249	0.99	1	0.17	0.23
13	10-49	0.92	1	0.4	0.38	41	10-49	1	1	0.21	0.21
13	250+	0.94	1	0.69	0.65	41	250+	0.9	1	0.71	0.5
13	50-249	0.99	1	0.32	0.31	41	50-249	0.99	1	0.1	0.11
14	10-49	0.85	1	0.38	0.41	42	10-49	0.99	1	0.11	0.12
14	250+	1	1	0.72	0.71	42	250+	1	1	0.26	0.24
14	50-249	1	1	0.23	0.23	42	50-249	0.98	1	0.08	0.1
15	10-49	0.97	1	0.23	0.22	43	10-49	0.98	1	0.24	0.24
15	250+	0.87	1	0.64	0.56	43	250+	0.68	1	0.14	0.4
15	50-249	0.98	1	0.42	0.39	43	50-249	0.99	1	0.17	0.17
16	10-49	0.98	1	0.17	0.19	45	10-49	1	1	0.1	0.1
16	250+	0.85	1	0.67	0.34	45	50+	0.9	1	0.1	0.23
16	50-249	0.92	1	0.3	0.34	46	10-49	1	1	0.11	0.11
17	10-49	1	1	0.15	0.15	46	250+	1	1	0.21	0.21
17	250+	0.78	1	0.29	0.24	46	50-249	1	1	-0.03	-0.03
17	50-249	0.96	1	0.12	0.12	47	10-49	0.96	1	0.08	0.08
18	10-49	0.9	1	0.11	0.1	47	250+	0.99	1	0.31	0.3
18	250+	0.82	1	0.12	-0.01	47	50-249	0.97	1	-0.02	-0.02
18	50-249	0.95	1	0.18	0.11	49	10-49	1	1	0.14	0.14
19	10-49	0.93	1	0.08	0.1	49	250+	0.71	0.99	0.3	0.43
19	250+	0.89	1	0.69	0.35	49	50-249	0.99	1	0.38	0.39
19	50-249	0.71	1	-0.02	0.53	50	10-49	0.72	1	-0.09	0.07
20	10-49	0.95	1	0.27	0.29	50	250+	0.93	1	0.69	0.52
20	250+	0.86	0.95	0.79	0.83	50	50-249	0.81	1	-0.1	-0.12
20	50-249	0.69	1	0.63	0.29	51	10-49	0.96	1	0.24	0.13
21	10-49	0.8	1	0.13	0.1	51	250+	0.66	1	-0.16	0.45
21	250+	1	1	0.22	0.23	51	50-249	NA	NA	NA	NA
21	50-249	0.99	1	0.38	0.4	52	10-49	0.99	1	0.29	0.27
22	10-49	1	1	0.39	0.39	52	250+	0.91	1	0.38	0.38
22	250+	0.76	1	0.15	0.48	52	50-249	0.96	1	0.52	0.39
22	50-249	1	1	0.23	0.22	53	ALL	1	1	1	1
23	10-49	0.97	1	0.16	0.15	55	10-49	0.86	1	0.06	0.07
23	250+	0.96	1	0.45	0.48	55	250+	1	1	0.07	0.08
23	50-249	0.99	1	0.28	0.28	55	50-249	0.96	1	0.26	0.39
24	10-49	0.99	1	0.36	0.31	56	10-49	0.99	1	0.11	0.12

24	250+	0.91	0.97	0.8	0.66	56	250+	0.86	0.97	0.87	0.53
24	50-249	0.99	1	0.14	0.13	56	50-249	0.95	1	-0.01	-0.02
25	10-49	0.86	1	0.16	0.22	58	10-49	0.93	1	-0.07	-0.07
25	250+	1	1	0.46	0.44	58	250+	0.9	1	0.6	0.73
25	50-249	0.97	1	0.3	0.33	58	50-249	0.97	1	0.06	0.01
26	10-49	0.99	1	0.47	0.49	59	10-49	0.92	1	0	-0.03
26	250+	0.92	0.99	0.86	0.72	59	250+	NA	NA	NA	NA
26	50-249	0.88	1	0.82	0.69	59	50-249	0.67	1	-0.05	-0.1
27	10-49	0.93	1	0.41	0.42	61	10-49	0.86	1	0.03	0.36
27	250+	0.92	0.99	0.8	0.65	61	250+	0.81	0.99	0.92	0.58
27	50-249	0.99	1	0.54	0.5	61	50-249	0.84	1	0	-0.09
28	10-49	0.9	1	0.14	0.21	62	10-49	0.95	1	0.2	0.25
28	250+	0.99	1	0.87	0.86	62	250+	0.91	0.97	0.88	0.67
28	50-249	1	1	0.25	0.24	62	50-249	0.89	1	0.05	0.07
29	10-49	0.75	1	0.27	0.28	63	10-49	1	1	0.25	0.25
29	250+	0.59	0.95	0.16	0.44	63	250+	0.94	1	0.38	0.38
29	50-249	0.96	1	0.27	0.35	63	50-249	0.99	1	0.11	0.12
30	10-49	0.92	1	0.07	0.37	64	10-49	1	1	0.11	0.12
30	250+	0.88	0.99	0.75	0.62	64	250+	0.9	0.99	0.39	0.53
30	50-249	0.99	1	0.53	0.5	64	50-249	1	1	0.43	0.43
31	10-49	0.99	1	0.24	0.24	65	10-49	1	1	0.5	0.51
31	250+	0.98	1	0.36	0.45	65	50+	0.97	1	0.27	0.23
31	50-249	0.98	1	0.41	0.48	66	10-49	1	1	0.37	0.36
32	10-49	0.88	1	0.53	0.39	66	50+	0.93	0.99	0.72	0.6
32	250+	0.89	1	0	0.06	68	10-49	0.98	1	0.1	0.13
32	50-249	0.97	1	0.21	0.31	68	50+	0.89	1	0.19	0.23
33	10-49	1	1	0.11	0.11	71	10-49	0.94	1	0.1	0.14
33	250+	0.74	1	0.9	0.46	71	250+	0.83	0.97	-0.04	0.18
33	50-249	1	1	-0.05	-0.04	71	50-249	0.98	1	0.1	0.1
35	10-49	1	1	-0.02	-0.02	72	10-49	0.99	1	0.7	0.65
35	250+	0.99	1	0.25	0.26	72	50+	0.74	0.92	0.14	0.54
35	50-249	0.98	1	-0.02	-0.03	77	10-49	0.98	1	0.15	0.17
36	10-49	0.66	1	0.85	0.48	77	50+	1	1	0.06	0.06
36	250+	0.81	1	0.59	0.53	B	10-49	0.95	1	0.26	0.31
						B	50+	0.56	1	1	0.55

4.4. Variations with respect to the first year of the reference period

The impact of the perturbation applied to turn06 is described by means of the ratio between turn06 and turn08. Table 5 presents the comparison between some statistical indicators assessed on original and perturbed ratios turn06/turn08. Only the combinations of nace and EMP08 that were changed with respect to these criteria are listed. The other 83% of combinations was unchanged from the point of view of the distribution of turn06/turn08.

Table 5. Assessment of the *turn06/turn08* perturbation (rounded values).

nace	EMP08	NbObs	MeanPerturbed/ MeanOrig	Variance Perturbed/Varian ceOrig	Corr(Per turbed,O rig)
11-12	250+	11	0.97	1.02	0.83
13	250+	33	0.98	1.43	0.89

15	250+	10	0.99	0.53	0.80
16	250+	9	1.00	0.71	0.79
18	250+	11	0.97	0.97	0.77
20	250+	41	1.00	1.00	1.00
22	250+	40	1.00	1.00	0.98
24	250+	42	0.99	1.00	1.00
26	250+	28	1.00	1.01	1.00
28	250+	121	1.00	1.04	0.99
31	250+	19	1.00	0.98	0.95
32	250+	12	1.00	0.96	0.96
33	250+	9	0.90	1.37	0.76
41	250+	21	1.02	1.81	0.88
43	250+	12	0.98	1.10	0.90
45	50+	130	1.00	1.01	1.00
50	250+	22	0.97	1.21	0.93
51	250+	6	0.96	0.92	0.82
53	ALL	21	1.37	19.07	0.17
55	250+	21	1.00	1.00	0.98
56	250+	41	0.99	1.03	1.00
63	250+	12	0.99	1.02	0.99
65	50+	66	1.00	1.08	0.98
66	50+	34	1.00	1.00	1.00
71	250+	15	0.97	1.16	0.86

4.5. Ratios rtot/turn08 and the other components of rtot

The data utility was assessed with respect to *rtot* and some of its components, for example, *rrdinx, rrdexx, rmacx*. Only the ratios *rtot/turn08*, *rrdinx/turn08, rrdexx/turn08, rmacx/turn08* were taken into account. For the Italian CIS2008 survey data, the quantiles of the above mentioned ratios, computed on the original and on the perturbed data, were compared. Generally, a very good agreement was observed. The maximum absolute difference between the quantiles of the ratios computed using the original values and quantiles of the ratios computed with perturbed values was 1.9, but in most cases the absolute difference between the quantiles of the ratios was inferior to 0.5.

5. References

1. Ichim, D. (2007), "Microdata Anonymisation of the Community Innovation Survey Data: a Density Based Clustering Approach for Risk Assessment", *Documenti Istat*, 2, available at www.istat.it.

2. Franconi, L. and Ichim, D. (2007), "Community Innovation Survey: comparable dissemination", Joint UNECE – Eurostat Worksession on Statistical Dislcosure Control,

Manchester, Manchester, United Kingdom, 2007, available at http://www.unece.org/stats/documents/2007.12.confidentiality.htm.

3. Ichim, D. (2009), "Disclosure Control for Business Microdata: a Density-Based Approach", *International Statistical Review*, (2009), 77, 2, 196–211.

4. Ichim, D. (2008), "Community Innovation Survey: a Flexible Approach to the Dissemination of Microdata Files for Research", European Conference on Quality in Official Statistics, available at http://q2008.istat.it/.

5. ISTAT (2008), "Statistiche sull'innovazione delle imprese", Collana *Informazioni*, N. 1.

6. Contacts

For questions on the survey, contact mastrost@istat.it.

For questions on the statistical disclosure control methodology, contact rilascio.microdati@istat.it.

7. Authors

This document was prepared by Daniela Ichim.