

*The challenge of a mixed-mode design survey and new IT
tools application: the case of the Italian Structure
Earning Surveys*

Fabiana Rocci
Stefania Cardinleschi
Stefanio De Santis

ISTAT

Presentation

- ❑ Description of the survey SES
- ❑ Introduction to the re-engineering of the flow
- ❑ Focus on the direct survey flow:
 - ❑ Editing and Correction process
 - ❑ Selective editing
 - ❑ imputation troigu administrative data
- ❑ Results of the direct survey



Description of the survey SES

- ❑ The SES final aim is to gather micro-data on employees about their individual and economic characteristics, to obtain detailed and comparable information on relationships between the level of remuneration and individual characteristics of employees and their employer
- ❑ The observation unit is the enterprise, through which the data are obtained on a representative sample of their employees
- ❑ The coverage is the economic activities from B to S of NACE Rev. 2, on business with 10 or more employees (the section O and units with less than 10 employees are optional)
- ❑ The collected data on the employees are:
sex, age, occupation (ISCO08), education (ISCED 97), full time/part time, mean monthly gross earnings in the reference month, mean annual gross earnings, number of hours paid includes all normal and overtime hours worked and remunerated by the employer during the reference month
- ❑ It is based on a two stage sampling design: a sample of employees nested in a sample of enterprises. The size of the second stage sample varies according to the enterprise size, from 10 – 19 employees a module for every employee, till 200 modules required to the biggest size-class (more than 10000 employees).



Description of the survey SES

Starting frame:

- ❑ Private sector: the starting frame is the official register of active enterprises defined by Istat (Asia), with several information about every unit as number of employees. From this frame a random stratified sample is drawn on business with 10 till 249 employees, it has been stratified according to the size class, Nace Rev. 2 groups and macro geographical area (NUTS1). On the other side a census is done on units with more than 250 employees
- ❑ Public sector (school excluded): the starting frame is the official Istat list of the Institutional units belonging to the PA sector, it is supplied with several information about every units but the number of employees. On this list a census is drawn
- ❑ Public School: the universe representing the starting frame has been elaborated on the basis of an integration of administrative data sources. Hence, the sample design of employers and employees belonging to the public Education sector has been realized through the integration of statistical archives provided by LFS and EU-SILC as well as cross linking techniques with administrative and fiscal data.



Re-engineering of the flow

The process has been re-engineered in order to improve the results in terms of:

- ❑ coverage
- ❑ response rate
- ❑ quality of the final estimates



Re-engineering of the flow

Taking into account

- ✓ the complexity of the two sample design
- ✓ variety of the starting frame

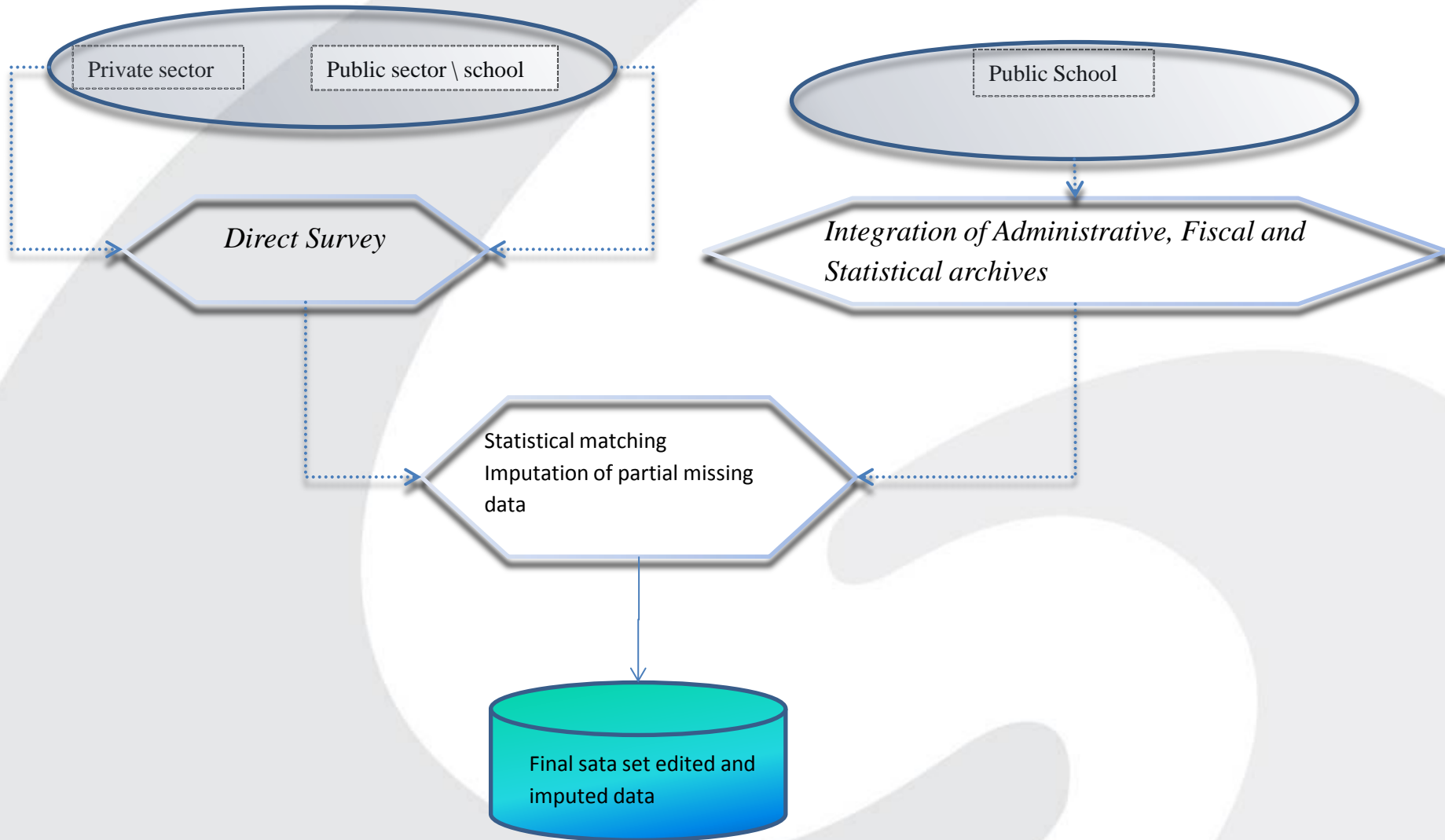
new technologies and methodologies have been considered, each of them to be suited to the given problem

- ❑ a computer assisted
- ❑ administrative data used as auxiliary information for the imputation of total non response (in critical strata)
- ❑ Integration of administrative, fiscal and statistical archives to estimate a complete sector

all this mixed with the different frames' characteristics made up two distinct flow at first that integrate at the last step

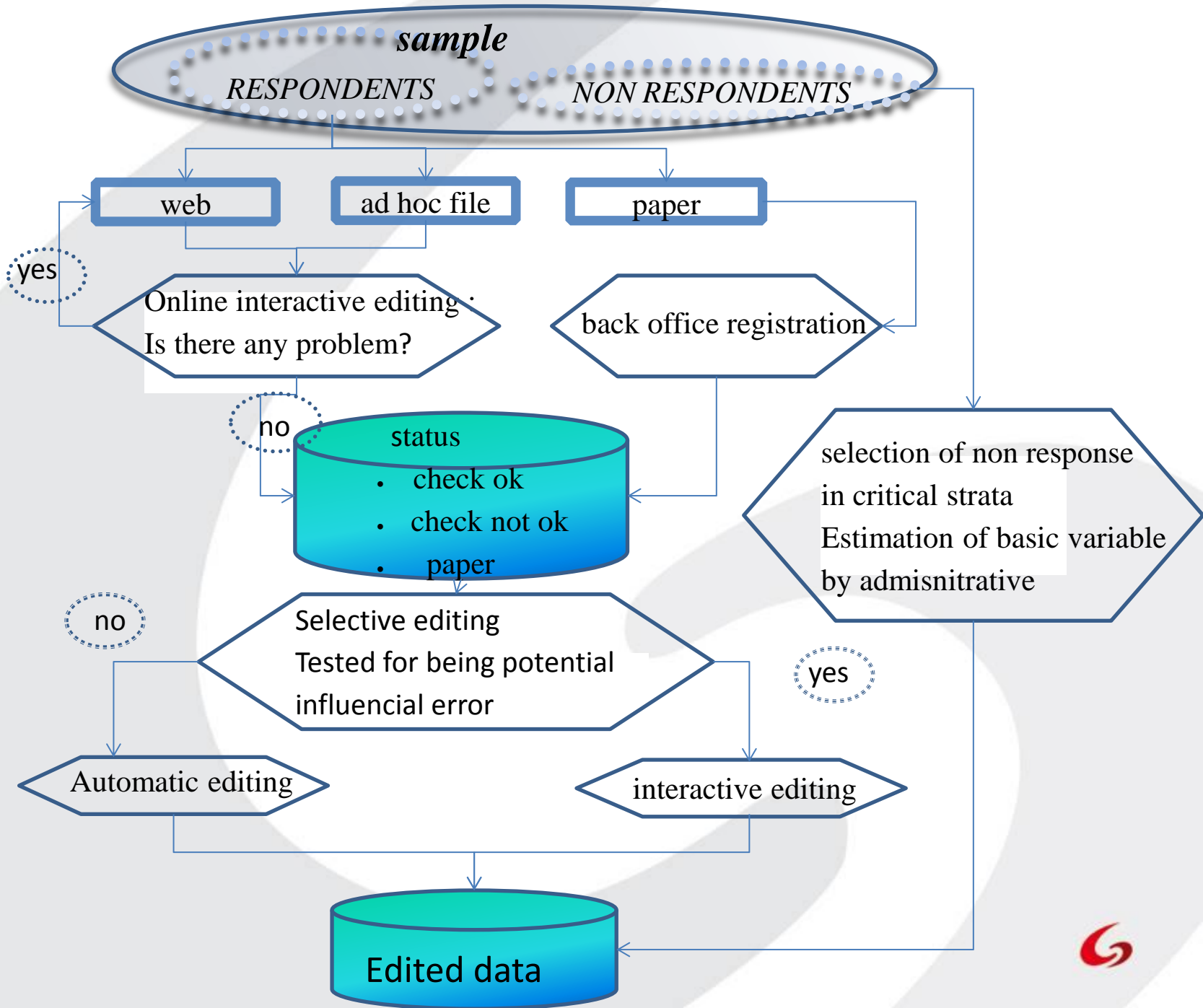


Re-engineering of the flow



D
i
r
e
c
t

S
u
r
v
e
y



Direct survey: coverage

Questionnaire by sector and mode

type	paper	ad hoc file	web	tot	starting sample
private	321	120	7587	8028	19534
public	5	10	558	573	1118
Tot	326	130	8145	8601	20652

42%

51%

	enterprises			employees		
sector	paper	ad hoc file	web	paper	ad hoc file	web
priv	4	1.5	94.5	2.6	3.6	93.8
pubb	0.9	1.7	97.4	0.7	3.3	96

in most cases the full answer were gain about the sample of employees resulting in a very huge amount of microdata : about 240.000, compared to the 160.000 of the previous session of the survey



final coverage by imputaion

Imputed enterprises by administrative data by size class and geographical area

size	north west	north east	centre	south	insular	tot
10-19	-	-	-	12	1	13
20-49	1	14	20	31	18	84
50-99	2	9	27	21	34	93
100-249	2	11	8	13	19	53
250-999	1	3	4	11	3	22
1000+	-	-	-	-	1	1
tot	6	37	59	88	76	266



E&I process: first step – selective editing

selective editing: to detect influential errors

i.e. to search for those units particularly abnormal and dangerous, according to the values of some variables of interest, which may affect the final estimates

selected enterprise by size

size	selected enterprises	imputed enterprises	labour cost variable	worked hours variable
10-19	7	7	4	3
20-49	4	4	2	3
50-99	-	-	-	
100-249	-	-	-	
250-999	41	28	11	20
1000+	30	16	9	11



Direct survey and E&I:

quality indicators of the raw data and analysis by mode

- ❑ To evaluate the efficiency of the interactive editing developed to be used on the web questionnaire
- ❑ To asses whether the mode used implies different quality of data



Frequency table of mode by...

enterprises by size and mode

size	paper	ad hoc file	web	tot
10-19	167	9	2369	2545
20-49	90	1	1405	1496
50-99	48	3	867	918
100-249	11	8	958	977
250-999	8	38	1878	1924
1000+	2	71	668	741
tot	326	130	8145	8601

enterprises by economic activity and mode

economic activity	paper	ad hoc file	web	tot
B	1	-	67	68
C	109	13	2607	2729
D	3	8	105	116
E	10	-	315	325
F	15	3	336	354
G	22	21	613	656
H	23	10	467	500
I	11	3	249	263
J	9	9	455	473
K	12	36	445	493
L	4	1	67	72
M	27	10	582	619
N	31	5	587	623
P	7	2	184	193
Q	14	9	607	630
R	18	-	300	318
S	10	-	159	169
tot	326	130	8145	8601

enterprises by geographical area and mode

geographical area	paper	ad hoc file	web	tot
north west	88	49	2740	2877
north east	80	27	2168	2275
centre	62	47	1626	1735
south	61	4	928	993
island	35	3	683	721
tot	326	130	8145	8601



Chi squared on ...

Test of Chi squared has been run, in order to test whether there is association between the chosen mode and given characteristics. For each of the previous table the test has been calculated and the results are in the following table:

	model	Test value	prob
mod 1	economic activity*mode	206.9007	<.0001
mod 2	geographical area*mode	54.4406	<.0001
mod 3	size *mode	558.6648	<.0001

Because the null hypothesis is lack of association, the diagnostic give us results to reject it in any model has been tested.



Quality of data according the mode

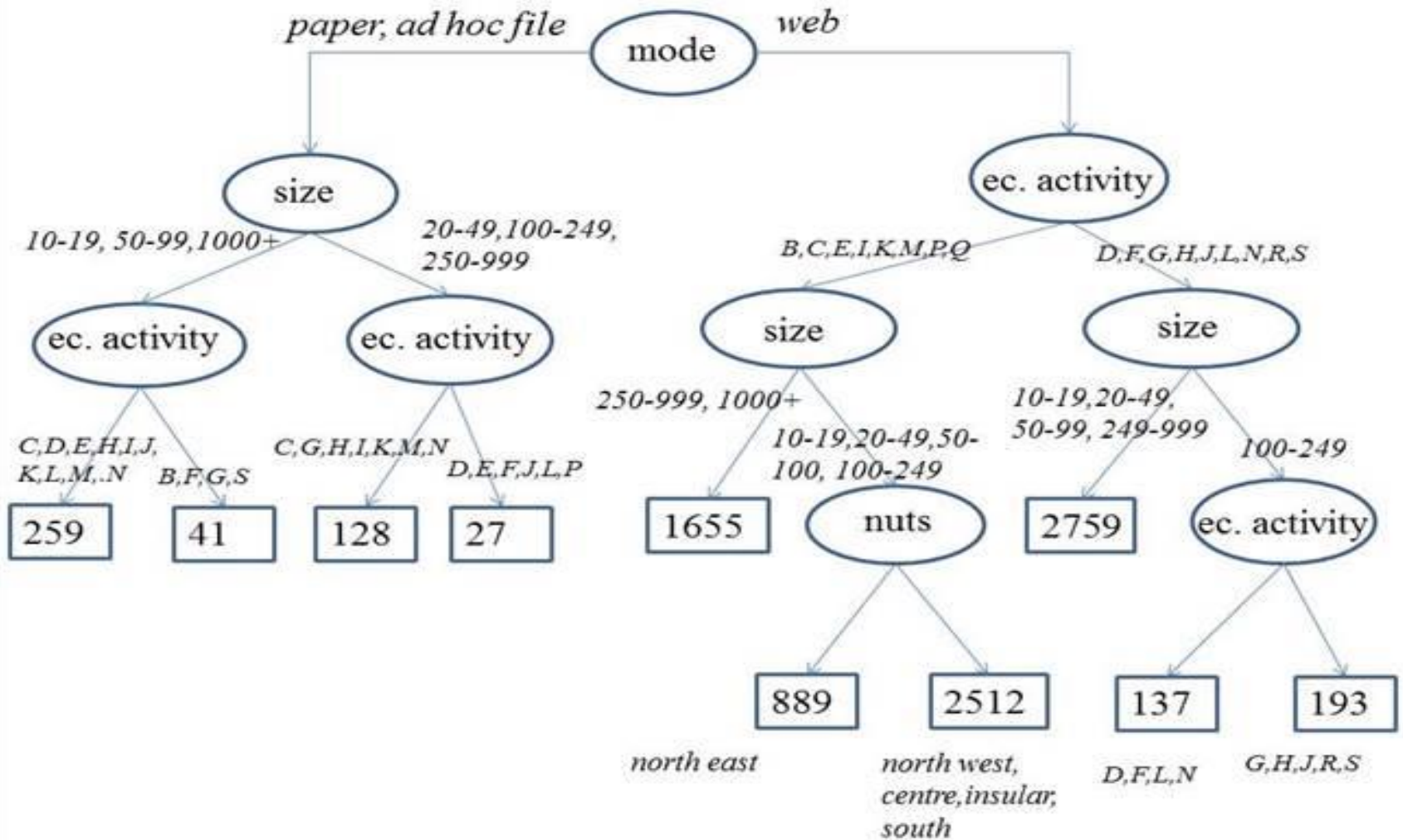
Synthetic table of percentage of fails by mode

17 edits out of the number 87, the most common

type of flag		paper	ad hoc file	web
Sum of type of employees	Total	10.5	6.3	7.4
	women	7.4	8.5	9.5
	partime	9.9	8.4	9.5
sum of worked hours by type of employees	total	9.9	9.2	9.4
	women	6.8	9.4	9.4
	partime	8.9	9.4	9.4
sum of wage component	total wage	10.2	4.2	6.3
check of partial missing data on worked hours by type of employees	manager	1	3.3	1.8
	white collar	0.6	1.4	1.7
	workers	1.3	0.6	0.8
coherence between variables	taxes as percentage of the total amountof labour cost	0.3	0.5	0.4
	hours worked per capita	7.2	10.2	9.5
	per capita labour costs	11.2	10.2	9.4
	limit of wage	11.9	10	9.1



Classification regression tree



conclusion

1. to extend the use of administrative data in order to fully integrate them with survey data only were necessary, it means for the not available data from the administrative sources
2. to optimize the use of IT tools for the direct survey with the web questionnaire
3. to optimize the use of selective editing also for to her variable, using as auxiliary variable also the ones from the administrative data
4. in the case there the will to give possibility to choose between different modes, the information about how it interact with the enterprise characteristics can be used in a very innovative way to better study a proper sample

