Divide, Allocate et Impera: Comparing Allocation Strategies via Simulation

Paola Maddalena Chiodini   Giancarlo Manzi   Bianca Maria Martelli
Flavio Verrecchia

Working Paper no. 9/2017

May 2017

UNIVERSITÀ DEGLI STUDI DI MILANO

Dipartimento di Economia, Management e Metodi Quantitativi

Via Conservatorio 7
20122 Milano
tel. ++39 02 503 21501 (21522) - fax ++39 02 503 21450 (21505)
http://www.economia.unimi.it
E Mail: dipeco@unimi.it

Divide, *allocate et impera*: comparing allocation strategies via simulation

Paola Maddalena Chiodini, Giancarlo Manzi, Bianca Maria Martelli, Flavio Verrecchia[1]

Abstract

In stratified sampling, the problem of optimally allocating the sample size is of primary importance, especially when reliable estimates are required both for the overall population and for subdomains. To this purpose, in this paper we compare multiple standard allocation mechanisms. In particular, standard allocation methods are compared with an allocation method that has been recently adopted by the Italian National Statistical Institute: the Robust Optimal Allocation with Uniform Stratum Threshold (ROAUST) method. Standard allocation methods considered in this comparison are: (i) the optimal Neyman allocation, (ii) the multivariate Neyman allocation, (iii) the Costa allocation, (iv) the Bankier allocation, and (v) the Interior Point Non Linear Programming (IPNLP) allocation. Results show that the optimal Neyman allocation method outperforms the ROAUST method at the overall sample level, whereas the latter method performs better at the stratum level. Some results on the Nonlinear Programming method are particularly interesting.

Keywords: Firm, Stratified Sampling, Permanent Random Numbers, Monte Carlo Simulation; Compromise Allocation; Interior Point Non Linear Programming.

Jel Classification: C83, C15, C81

[1] Paola M. Chiodini, Department of Statistics and Quantitative Methods, University of Milano-Bicocca, Italy. E-mail: paola.chiodini@unimib.it.
Giancarlo Manzi, Department of Economics, Management and Quantitative Methods, Università degli Studi di Milano, Italy. E-mail: giancarlo.manzi@unimi.it.
Bianca M. Martelli, Local Office for Marche, Tuscany, and Umbria, National Institute of Statistics, Italy. E-mail: biancamaria.martelli@gmail.com.
Flavio Verrecchia (*Corresponding Author*), North-West Region Office, National Institute of Statistics, Italy. E-mail: verrecchia@istat.it.

Any completed sample is potentially a guide to improved future sampling
(W. G. Cochran (1977), Sampling techniques, Wiley, New York, p.7)

1. Introduction

The last update of the NACE - General Classification of Economic Activities within the European Communities (Rev. 2.2 - European Commission 2006b; Eurostat, 2008) - led many European national statistical agencies to revise their standard procedures in business survey sampling. Questionnaires, survey designs, and methods of estimation needed a revision in order to maintain the comparability with the previous NACE classification and the integrity of time series. In Italy, the Italian National Institute of Statistics (ISTAT) also updated its business survey designs and estimation methods as a consequence of these changes. In particular, the Italian Business and Consumer Survey (BCS) on Manufacturing needed a radical revision. Since the economic activities used for stratification in this survey changed in NACE Rev. 2, a validation of previously used methodologies and procedures was needed. This paper is focused on testing and validating the strata allocation methodology to overcome the NACE Rev. 2 changes. Then, keeping in mind that optimal strata allocation should retain multiple sources of information and produce reliable estimates, we compare some popular allocation methods with a new method, the Robust Optimal Allocation with Uniform Stratum Threshold (ROAUST), which is based on an “allocation compromise”. A Monte Carlo simulation (MC-simulation) device based on a single-step experiment is used for simultaneous validation (Chiodini *et al*., 2008; Chiodini *et al*., 2011).

The classical optimal Neyman allocation for stratified sampling (Neyman, 1934) is a widely applied sampling design in business surveys, used mainly for collecting quantitative information (Smith *et al*., 2003). This popularity stems from the capability of the samples generated through this allocation to retain information and efficiency in challenging sampling difficulties, for example, when the sample size is small and a given (high) level of precision is required. Its implementation is relatively simple and usually provides efficient estimators for population parameters (Kozak *et al*., 2007; Särndal *et al*., 2003). However, when a suitable compromise between budget constraints and a given level of estimate precision cannot be reached, the resulting allocation may be sub-optimal (James *et al*., 2005). Moreover, it could be the case that there exists a trade-off between the optimal allocation principle and the need of a minimum information from each stratum.

This paper belongs to the field of sampling allocation when reliable estimates are required both at national and subnational level. In this framework, many authors have proposed in recent years some satisfactory and more flexible solutions. For example, Bankier (1988) suggests a ‘power allocation formula’ to be used in designing several official surveys at Statistics Canada. This power allocation formula is essentially a compromise allocation, which includes the Neyman allocation. Costa *et al*. (2004) propose a compromise allocation based on a combination of the proportional and the uniform allocation. Choudry *et al*. (2012) propose a non-linear programming method to obtain an optimal sample stratum allocation that minimizes the total sample size, conditioned on a fixed coefficient of variation for the estimators of the stratum means and the population mean. In both Bankier’s and Choudry *et al.’*s approach the focus is on the stratum coefficient of variation. In Choudry *et al.’s* several aspects showing computational efficiency and originality emerge. Among other things, their approach has the quality of using a specific domain and population tolerances for the coefficient of variation. Contrary to these proposals, the ROAUST method is easy to implement in applied contexts, and allows for the *ex-ante* setting of the uniform stratum fraction for a subsequent calculation of the coefficient of variation and the Neyman allocation quotas, even during the survey implementation. Costa *et al.*’s proposal is therefore similar to ROAUST in that both methods search for a family of compromise allocations, which from one side aim at finding sufficiently informative strata (the uniform part). However, from another side, Costa *et al*.’s proposal takes into account also the sample proportionality, but leaves apart the allocation optimality considered in the ROAUST approach.

Differently from other proposals, the ROAUST method allows for deriving a general formulation, which brings in multiple ways of performing stratified sampling in the daily work of several statistical agencies under the same hat. Even though working reports and documents are not easily available in the literature, and, in any case, difficult to retrieve on the World Wide Web, many statistical offices have experimented some sort of “sample size constraining” with applied purposes. For example, James *et al*. (2005), while working at the UK Office for National Statistics (ONS), and Buglielli *et al*. (2011), while working at ISTAT, have recently used constrained sampling allocation. In particular, as reported in ONS survey accompanying documents, this constrained allocation is used: (i) to overcome the issue of non-response; (ii) to help reducing the potential bias in the estimates by ensuring a minimum sample size. In the ISTAT case, the MAUSS software package (Istat, 2004) forces a minimum number of units per stratum equal to two when calculating the indicators of stratum variability. This can be set also in the ROAUST approach.

Stratification is often based on predetermined rules according to some fixed administrative settings (e.g. fixed NUTS areas; European Union, 2003) and economic classifications (e.g. firm sizes and sectors of activity). Therefore, it is not always possible to adjust *ad libitum* the stratification process to increase the stratum homogeneity so as to optimize a survey plan. Secondly, in business surveys the firm size tends to have positively skewed distributions (Smith *et al*., 2003).

In the BCS, the frame is selected by applying a cut-off, which excludes firms with fewer than 10 persons employed (Malgarini *et al*., 2005). It is worth noticing that several empirical analyses performed in the past have shown that some issues like the non-response rate, the birth-mortality rate of economic units and the errors of under- or over-coverage of the frame become more relevant as the size of the units gets smaller (Cochran, 1977; ISTAT, 1989), which is a feature characterising the Italian industrial system in recent years, as the 2001-2011 Italian Censuses outline, and data from the Italian Business Register clearly show (ISTAT 2005, ISTAT 2005-2011a, ISTAT 2005-2011b).

This paper is organised as follows. In Section 2, the ISTAT BCS on manufacturing is briefly introduced. In Section 3 selected strata and frame used are presented; in Section 4 selected allocation methods are explained, and in Section 5 a simulation procedure to test the performance of the allocation methods is described. Section 6 contains the simulation results. Section 7 comments and conclusions are presented.

2. The Italian BCS on Manufacturing

In 2011, ISTAT took over the BCS in the Italian manufacturing industry formerly conducted by ISAE (Istituto di Studi e Analisi Economiche) and ISCO (Istituto per lo Studio della COngiuntura) since 1959

(ISCO, 1961; Pinca, 1990; Martelli, 1998) in the frame of the European Commission (EC) Harmonised Project (European Commission 2006a, 2014).

From its origin in 1961 (EC, 1961), the survey is part of the Joint Harmonised BSC Program of EC, subsequently regulated by an EC Decision (EC, 1997) and an EC Communication (EC, 2006c). The project has continued over the years according to the European guidelines, and the adopted sampling techniques and design were continuously upgraded also according to OECD (2003) guidelines (Malgarini *et al*., 2005). The survey is conducted in each country according to a shared methodology, which essentially consists of a harmonised questionnaire and some common features (high frequency, timeliness in collection and dissemination of results and continuous harmonisation) (EC, 2014). Apart from a given sample size of about 4,000 units recommended by the EC, ISTAT is however relatively free to define any other aspects of the entire sampling process.

Although initially conceived as a purposive panel among managers ("expert witnesses"), the BCS sample on Italian manufacturing has been developed over the years with the aim of better matching the methodological advances in sampling theory.

Since 1998, also in accordance to the recommendations of international organizations (Martelli, 1998; EC, 2006a, 2014; OECD, 2003), the BCS has adopted a stratified sampling design with univariate Neyman x-optimal allocation (see Section 3 below).

The survey investigates the "confidence" of economic agents, that can be defined as the set of attitudes on which behaviours and economic decisions (on production, investment, consumption, etc.) are based (Katona, 1975), by asking entrepreneurs and managers about assessments on current trends and expectations for the near future on both their own business and the general situation of the economy (UNSD, 2013). It could be worth giving a concise explanation of how the confidence indicator is calculated and applied, as confidence series and indicators are widely used in cyclical analysis even though there is no clear economic theory explaining why this kind of information is suitable to monitor the economic fluctuations (Zarnowitz, 1992). As data are mainly categorical, for each firm and each question, within a three-level Likert scale (i.e. "high", "normal" and "low"), the collected answers are recorded assigning value 1 to the selected reply option and 0 otherwise. For example, if the question is: "Excluding seasonal variations, do you consider that in volume terms your present order book is: “Above normal”; “Normal”; “Below normal”; and the respondent chooses “Normal”, then for this question the code string will be “010” (dropping the “not available” replies). Data are then processed with a double quantitative weighting system in order to transform them in percentages and obtain “balances”, that is monthly quantitative values for each question in each stratum of the sample. “Weighting can be seen as the first step of the quantification process applied to categorical data to obtain confidence quantitative estimates” (UNSD 2013, par 5.9). The various steps are synthetically reported in Table 2.1. These balances are also called confidence series of the investigated phenomena (e.g. order books, production expectations, etc.). The industrial composite confidence indicator, which allows reducing the risk of false signals, is then calculated according to the classical NBER approach as simple arithmetic mean of balances of order level, production expectation and stocks level (with inverted sign) (EC, 2006a, 2014). Confidence and composite confidence indexes show satisfactory positive correlation with the cyclical component (suitably extracted) of the corresponding reference quantitative series (see e.g. OECD, 2003 sec. 7, where examples of correlations ranging between 0.76 to 0.87 are presented). These series are therefore widely used for timely information on the evolution of economic phenomena. From a statistical point of view these correlations also validate the sample design, as it is not possible to compare confidence outcomes with corresponding population values.

Table 2.1
Main features of the Italian Confidence Business Survey on Manufacturing

Starting date	1961
Timing	Monthly, fieldwork in the first half of the reference month; dissemination of results within the end of the reference month
Frame	Italian Active Business Integrated Statistical Register - A.S.I.A. (Archivio Statistico Imprese Attive). Each year t the archive is updated with reference to year *t*-2
Target population	Manufacturing businesses with more than 10 persons employed (86,447 units in 2007)
Sample unit	Firm
Variables/questions	Qualitative questions mainly on a 3 points Likert scale according to the Harmonised EC Questionnaire (7 monthly and 9 quarterly): assessments and/or expectations on order book, production, stocks, prices, employment, etc.
Data Collection Mode	Mixed, mainly CATI and some Fax
Sample design	Stratified random sample (by 4 NUTS1 geographical partitions; 19 NACE Rev. 2 two-digit level economic sectors -from 10 to 33- and three size classes: 10-49, 50-249, more than 249). 228 strata (2 empty) in 2007 update. Panel sample with substitution of non respondent units only

Unit allocation	Univariate Neyman x-optimal allocation (with x=persons employed)
Sample size / Coverage / Sampling error	About 4,000 monthly interviews. Coverage: 6% of population firms; 9% of population persons employed. Sampling error: lower than 4.4%
Estimation/weighting	Data estimated in two steps: 1) applying *size weights* to each unit and each question: persons employed declared by firms to transform categorical replies in percentages according to relative importance of the firms; 2) applying stratum weight according to value added. Balances for each question calculated, for each stratum, as simple differences between favourable and unfavourable reply options. The series stemming from the procedure over the time are seasonally adjusted.

Source: EC BTS Metadata (Available at: http://ec.europa.eu/economy_finance/db_indicators/surveys/documents/metadata2/metadata_it_indu_en.pdf)

In Italy, the adopted sample structure has produced satisfactory results during the years, as the business confidence indicator shows to be a good proxy of the economic evolution of the country. A visual inspection of Figure 2.1 points out the capability of the survey to fit and sometimes foresee the economic cycle. In particular, as the survey results represent a sort of deviation from a ‘normal’ level of the economic evolution of the country, confidence could be used to monitor the so called ‘growth rate cycle’, that is the evolution of the (annual) growth rate of the reference variable (in this case the GDP). Such a growth rate is defined in annual difference because in this way cyclical movements are amplified, while, on the other hand, the first difference would dampen cyclical movements and only amplify the series irregularity. Furthermore, due to the availability of the outcomes at the end of each reference month, information about economic variables is obtained well ahead of the official statistics (for example, GDP estimates are available six months before the official release).

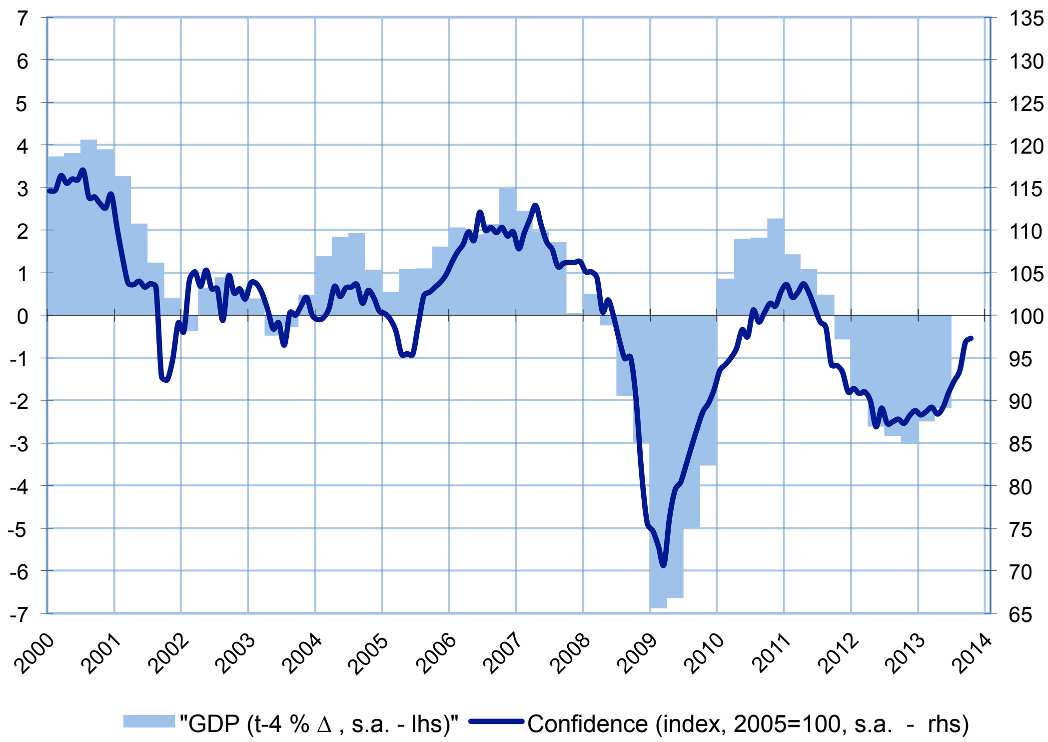

Figure 2.1 Italian GDP and Confidence

Source: ISTAT.
Notes: rhs: Right-Hand Side; lhs: Left-Hand Side; s.a.: seasonally adjusted; year t-4 Δ in percentage: GDP yearly percentage changes on quarterly data; Confidence: monthly values

3. Frame and Strata

Besides the preservation of a good quality of the estimates for the overall population, allocations methods also focus on improving the precision of the estimates *within* the strata. With this in mind, we discuss here the most important aspects to be analysed while performing a sampling plan like the one adopted in the BCS.

The available business register for Italy is the *Statistical Archive of the Active Business* (ASIA – ISTAT, 2007) provided by ISTAT, which is in line with the NACE Rev. 2 classification (ISTAT, 2009). Starting from the 2007 release, this register provides an exhaustive list of all the Italian firms. This is the frame we will use in our comparison.

As already mentioned, in the ASIA archive a lower cut-off is applied in selecting the frame by excluding firms with less than 10 persons employed (Elisson and Elvers, 2001). Therefore, the selected frame comprises details for just less than 90 thousands manufacturing firms (about 18% of all Italian manufacturing firms), accounting for almost 77% of the economic activity in terms of number of persons employed (Table 3.1).

In the European BCS, the strata are usually defined according to three variables: firm size (generally in terms of persons employed), economic classification and geographical areas. In the present analysis the firm size refers to 3 classes (according to the European Commission (EC) recommendations, EC 2003): small firms (10 - 49 persons employed), medium-sized firms (50 - 249 persons employed) and large firms (with at least 250 persons employed). The selection of these classes was specifically requested by the EC for harmonisation reasons. It is worth noticing that this requirement prevents us from using other possible selections of stratum boundaries (e.g. more suitable to fit the different regional and industrial features).

The economic sectors mainly reflect the NACE Rev. 2 two-digit classification with a few further groupings. The geographical classification relies on the Nomenclature of Territorial Units for Statistics (NUTS), the standard geocoding for referencing the subdivisions of countries for statistical purposes among the EU member states. With regard to Italy, defining the country as level 0, level 1 indicates the geographical partitions and level 2 the administrative regions. With respect to NUTS-2, the selection of NUTS-1 breakdown, which comprises only four partitions, allows to nearly completely avoid the occurrence of empty strata in the frame (data are missing only in strata “NACE sect. 19 / size 250+ / North-east” and “NACE sect. 20 – 21 / size 250+ / South”) and, although the within-variance increases, this occurrence does not have any substantial impact on sample designs with regards to optimality (Chiodini *et al.*, 2008). The sample design consists of 226 strata (i.e. 228 strata from the classification cardinality minus two empty strata). These stratum definitions essentially derive from administrative settings, and often do not respect the statistical and economic principles of stratum definition (see, e.g., Kozak and Jankowski, 2008). However, the involved entities (the EC, governmental institutions and economic operators) are deeply interested and greatly support (also financially) the production of detailed domain statistics. Furthermore this geographical stratification could be seen as a “primary” stratification, as it was applied in the past to get sufficiently large sample size also in relatively small regions (see Section 4.4 below).

However, from an international point of view, the NACE economic stratification plays a role of primary importance, as sectorial findings are also obtained. Finally, the stratification by firm size is also justified by economic and statistical needs for investigating and analysing different firms' behaviours. As a consequence, in this paper the focus is set, besides the whole sample results, directly on these elementary strata. Table 3.1 shows the applied frame by NUTS, sectors and firm size, respectively.

Table 3.1
Firms by stratum (NUTS-1, firm size, economic sectors), Italy, 2007 (units)

NUTS1 and firm size / NACE Sectors	North-West			North-East			Centre			South			Total
	10-49	50-249	250+	10-49	50-249	250+	10-49	50-249	250+	10-49	50-249	250+	
10-12.	1566	228	54	1782	278	39	1156	86	15	1978	177	15	7374
13.	1503	328	52	557	75	9	983	80	3	273	25	-	3890
14.	1398	128	22	1817	138	26	1217	99	8	1317	86	6	6262
15.	309	45	-	879	117	14	2095	136	9	628	46	3	4283
16-17.	1292	148	20	1524	163	14	898	84	10	780	48	4	4985
18.	952	79	11	739	58	5	507	34	-	298	19	-	2704
19.	31	10	7	17	5	.	21	7	4	74	6	4	186
20-21.	668	291	89	346	115	15	230	63	32	231	33	.	2113
22.	1607	288	41	1003	189	16	553	89	4	447	61	7	4305
23.	919	127	19	1199	229	50	878	103	16	1236	95	-	4872
24.	641	202	42	277	115	16	162	32	8	161	34	5	1695
25.	6428	635	42	4799	443	31	2074	177	9	2086	214	11	16949
26.	738	146	29	430	103	15	277	57	13	122	23	4	1957
27.	1060	204	36	842	162	28	353	60	15	198	23	3	2984
28.	3247	665	88	2823	608	105	750	128	8	521	57	4	9004
29-30.	571	183	79	373	99	31	333	71	15	245	87	19	2106
31.	898	90	5	1659	234	19	925	101	8	475	49	6	4469
32.	598	80	12	692	107	7	492	36	4	194	6	-	2229
33.	1408	87	5	958	43	3	674	27	4	802	63	6	4080
Total	25834	3964	655	22716	3281	443	14578	1470	186	12066	1152	102	86447

Source: Data processing on ISTAT, 2007 data
Notes: '.' Missing value; '-' less than 3 units; 10-12. Manufacture of food, beverages and tobacco products; 13. Manufacture of textiles; 14. Manufacture of wearing apparel; 15. Manufacture of leather and related products; 16-17. Manufacture of wood and paper products; 18. Printing and reproduction of recorded media; 19. Manufacture of coke and refined petroleum products; 20-21. Manufacture of chemical and pharmaceutical products; 22. Manufacture of rubber and plastic products; 23. Manufacture of other non-metallic mineral products; 24. Manufacture of basic metals; 25. Manufacture of fabricated metal products, except machinery and equipment; 26. Manufacture of computer, electronic and optical products; 27. Manufacture of electrical equipment; 28. Manufacture of machinery and equipment n.e.c.; 29-30. Manufacture of transport vehicles; 31. Manufacture of furniture; 32. Other manufacturing; 33. Repair and installation of machinery and equipment.

4. Allocation Methods

In this section allocation methods to be used in the simulation involving the whole sample and each stratum generated by the three stratification variables will be presented. The selection of these methods stems from the literature strengthening the validity of the Neyman allocation, which is considered optimal compared to more classical methods like the Uniform or the Proportional allocation and therefore will be the main benchmark in our simulation.

The variable *person employed* is widely used for variance estimation, as it is generally easily available for all population units (firms), and therefore suitable also for our exercise, being usually used for GDP estimation purposes. In simulation exercises we used the information available from register, instead of the survey itself, as it allowed us to perform comparisons between the different sampling design considered.

In our simulation, we produce samples of almost identical size for each allocation method. According to EC recommendations, the sample size ($n \approx 4000$ units) has to be predetermined. This bound state that gains in precision of estimates can be obtained only by choosing more efficient sample allocations.

In the next sub-sections the selected methods are briefly presented and discussed in their theoretical features. Algorithms implementing these methods to be used in our simulation are also presented, together with some adaptations needed for a suitable comparison. Sub-sections 4.1 to 4.3 deal with the classical methods used in this paper as benchmarks for other more complex allocation methods presented, and discussed in the following sub-sections 4.4 to 4.11. Among the considered benchmark allocations, it is worth noticing the exclusion of the random allocation, which was tested in Chiodini *et al.* (2011), showing a high volatility in the results therein presented.

4.1 The Uniform Allocation

The uniform allocation implies a constant stratum sample size n_h, which is set independently on the population strata size N_h. Let $n = \sum_h n_h$ be the overall sample size. If $N_h > 0$, the stratum sample size is given by:

$$n_h = \frac{n}{H}, \quad h=1,\ldots,H. \qquad [4.1]$$

The uniform allocation ensures a non-null sample size in each stratum also in those strata where the population size is very small, regardless proportionality criteria. In our case, given $n \approx 4000$ and $H=226$ and n/H is about 18. In the population of Italian firms presented in Table 3.1, 57 strata have a stratum size $N_h<18$. Therefore, we apply a slightly adjusted uniform allocation method in the simulation when $n/H > N_h$. This is performed as follows. Let $\mathcal{B}$ be the set of m strata where $n/H > N_h$, and $\mathcal{A}$ the set of strata where $n/H \leq N_h$. Compute:

$$n_h^* = \frac{n - \sum_{h \in \mathcal{B}} N_h}{H - m}. \quad [4.2]$$

Then, n_h becomes:

$$n_h = \begin{cases} n_h^*, & h \in \mathcal{A} \\ N_h, & h \in \mathcal{B} \end{cases}. \quad [4.3]$$

This modified uniform allocation can be formulated in algorithmic form as follows:

****

Algorithm 1.

Step 1. Set the array of population stratum sizes N_h $(N_1, \dots, N_H)$, the population size $N = \sum_{h=1}^{H} N_h$, the sample size $n = \sum_h n_h$, the initial uniform stratum sample size $n_h = \frac{n_0}{H}$.

Step 2. Initialize $\mathrm{n} = 0,\ n^{\mathcal{A}} = 0,\ n^{\mathcal{B}} = 0, n^{\mathrm{r}} = 0, m = 0, n_0 = 4000$ and $(n_1, \dots, n_H) = (0, \dots, 0)$.

Step 3. For *h*=1 to *H*

If $N_h < \lfloor n_h \rceil$ Then Do

Set $h \in \mathcal{B}$

$n_h^{\mathcal{B}} = N_h$

$n^{\mathcal{B}} = n^{\mathcal{B}} + n_h^{\mathcal{B}}$

$m = m + 1$

End

Next *h*

Step 4. $H^{\mathcal{A}} = H - m$; $n^{\mathcal{A}} = \lfloor n_h \rceil H^{\mathcal{A}}$; $n^r = n_0 - n^{\mathcal{B}} - n^{\mathcal{A}}$

Step 5. Do Until $\lfloor n^r / H^{\mathcal{A}} \rceil = 0$:

$n_h = n_h + \lfloor n^r / H^{\mathcal{A}} \rceil$

Do Step 2, Step 3, Step 4 End

End

Step 6. For *h*=1 to *H*

If $h \notin \mathcal{B}$ Then Do

Set $h \in$ A

End

Next *h*

Step 7. $n = n^{\mathcal{A}} + n^{\mathcal{B}}$; $n_h^{\mathcal{A}} = \frac{n^{\mathcal{A}}}{H^{\mathcal{A}}}$

In this algorithm n^r is the units not assigned to any stratum and $n_h^{\mathcal{A}}$ is the adjusted stratum sample size for those strata for which $N_h \geq n_h$. In our case two iterations are enough to reach the target sample size. In the first iteration the sample size for the equal allocation is *n*=3486 ($n^{\mathcal{B}} = 444$; $n^{\mathcal{A}} = 3042$) and the number of residual units is n^r=514. In the second iteration the sample size for the equal allocation is n=3984 ($n^{\mathcal{B}} = 540$; $n^{\mathcal{A}} = 3444$) and the number of residual units is n^r=16.

****

In our case we observe m=62 strata with stratum size N_h<21, for a total amount of $\sum_{h\in\mathcal{B}} N_h$=540 units involved. In the remaining $H-m$=164 strata, the sample stratum size is equal to 21, summing up to $n-\sum_{h\in\mathcal{B}} N_h$=3444 units involved. More precisely, in these $H-m$ strata the size sums up to $(n-\sum_{h\in\mathcal{B}} N_h)/\,(H-m)=21$, as n sample units are needed.

The uniform allocation has the advantage to avoid the occurrence of empty strata, and provides a disproportionate allocation, which gives a larger sampling fraction and therefore more weight to small strata.

4.2 The Proportional Allocation

The proportional allocation represents the simplest methodology to build a self-weighting sample, although it could be unsatisfactory at stratum level when the strata size is small.

In the proportional allocation the sample size n_h is given by:

$$n_h = nW_h \qquad h=1,\dots,H \qquad [4.4]$$

where $W_h = N_h/N$ is the relative stratum size.

In this paper a slightly modified proportional allocation method is applied due to the stratification design. Precisely formula 4.6 is modified applying a uniform stratum threshold equal to 1, for N_h>0, as follows:

$$n_h = 1 + (n-H)W_h \qquad h=1,\dots,H. \qquad [4.5]$$

This adjustment ensures a non-null sample size in each stratum also in those strata where the population size is very small (in our case n/N=4.63%; therefore in 41 strata we have $n_h < 0.5$).

This procedure can be formulated in an algorithmic form by first assigning a uniform threshold equal to 1, and then a proportional quota for those strata where $nW_h > 1$. The algorithm is the following:

****

Algorithm 2.

Step 1. Set the array of relative population stratum sizes W_h $(W_1, \dots, W_H)$, the sample size $n = \sum_h n_h$, the initial proportional stratum sample size $n_h = n_0 W_h$.

Step 2. Initialize $n = 0,\ n^{\mathcal{A}} = 0,\ n^{\mathcal{B}} = 0,\ n^{\mathrm{r}} = 0,\ n_0 = 4000$ and $(n_1, \dots, n_H) = (0, \dots, 0)$.

Step 3. For *h*=1 to *H*

If $N_h > 0$, then $n_h^{\mathcal{B}} = 1$ End

$n^{\mathcal{B}} = n^{\mathcal{B}} + n_h^{\mathcal{B}}$

Next *h*

Step 4. $n_0^{\mathcal{A}} = n_0 - n^{\mathcal{B}}$

Step 5. For *h*=1 to *H*

If $n_0 W_h > 1$, Then $n_h^{\mathcal{A}} = \lfloor n_0^{\mathcal{A}} W_h \rceil$ End

$n^{\mathcal{A}} = n^{\mathcal{A}} + n_h^{\mathcal{A}}$

Next *h*

Step 6. For *h*=1 To *H*

$n_h = n_h^{\mathcal{B}} + n_h^{\mathcal{A}}$

$n = n + n_h$

Next *h*

Step 7. $n^r = n_0 - n$

In this algorithm n^r is the number of units not assigned to any stratum and $n_h^{\mathcal{A}}$ is the adjusted stratum sample size for those strata for which $n_0 W_h > 1$. $n^{\mathcal{B}} = 226$, $n_0^{\mathcal{A}} = 3774$, $n^{\mathcal{A}} = 3760$. In this algorithm, the final resulting sample size is *n*=3984 ($n^{\mathcal{B}} = 226;\ n^{\mathcal{A}} = 3760$) and the number of residual units is n^r=16.

****

4.3 The Optimal Allocation

Given the target sample size $n = \sum_h n_h$, the optimal allocation method allows increasing the precision of the estimates by assigning different sampling fractions to the strata. This is performed by letting the sampling fractions be proportional to both the stratum standard deviation and the population size within each stratum (Kish, 1965).

Given that for this method it might be the case that $\frac{n_h}{N_h} > 1$, the classical Neyman strata allocation formula, has to be re-written (without taking into account the cost function, which will be considered constant in our simulation). Let

$$n_h^* = n\frac{N_h\sigma_h}{\sum_{h=1}^{H} N_h\sigma_h} \qquad [4.6]$$

be the Neyman allocation formula for stratum $h \in \mathcal{A}$. $\mathcal{A}$ is the set of strata, where $n_h^* < N_h$. $\mathcal{B}$ is the set of strata where $n_h^* \geq N_h$. Then set n_h as in formula [4.3].

In general, this procedure can be formulated in algorithmic form as follows:

****

Algorithm 3.

Step 1. Set $(\sigma_1, \dots, \sigma_H)$ as the array of population stratum standard deviation σ_h, $(N_1, \dots, N_H)$ as the array of population stratum sizes N_h, $N = \sum_{h=1}^{H} N_h$ as the population size, $n = \sum_h n_h$ as the sample size, $n_h = n_0 \frac{N_h\sigma_h}{\sum_{h=1}^{H} N_h\sigma_h}$ as the initial Neyman stratum sample size. Consider the following two macros:

Set_A: If $N_h \geq \lfloor n_h \rceil$ Then Do

Set $h \in \mathcal{A}$

$n^{\mathcal{A}} = n^{\mathcal{A}} + n_h$

End

Set_B: If $N_h < \lfloor n_h \rceil$ Then Do

Set $h \in \mathcal{B}$

$n_h = N_h$

$n^{\mathcal{B}} = n^{\mathcal{B}} + N_h$

$m = m + 1$

End

Step 2. Initialize $n = 0$, $n^{\mathcal{A}} = 0$, $n^{\mathcal{B}} = 0$, $n^{\mathrm{r}} = 0, m = 0$, $n_0 \leq 3984$ (i.e. the total sample size obtained in the uniform allocation), and $(n_1, \dots, n_H) = (0, \dots, 0)$.

Step 3. For *h*=1 To *H*

Call Set_A

Call Set_B

Next *h*

Step 4. $n^{\mathrm{r}} = n_0 - n^{\mathcal{A}} - n^{\mathcal{B}}$

Step 5. Do Until $\lfloor n^{\mathrm{r}} \rceil \leq 1$:

For $h \in \mathcal{A}$ Then Do:

$n_h = \left(n^{\mathcal{A}} + n^{\mathrm{r}}\right) \frac{N_h\sigma_h}{\sum_{h=1}^{H} N_h\sigma_h}$

Next *h*

Initialize $n^{\mathcal{A}} = 0$

For $h \in \mathcal{A}$ Then Do

Call Set_A

Call Set_B

Next h

Do Step 4 End

End

Step 6. $n = n^{\mathcal{A}} + n^{\mathcal{B}}$

In this algorithm n^r is the number of units not assigned to any stratum. After performing the algorithm, the final resulting sample size is n=3983 ($n^{\mathcal{B}} = 897$; $n^{\mathcal{A}} = 3086$), the number of residual units is n^r=1 (which derives from the uniform allocation) and $m = 34$.

****

4.4 The ISAE Constrained Allocation

The allocation method applied by ISAE is simply a slight modification of the classical x-optimal Neyman criterion applied separately to each of the k, for k=1,...,K, NUTS-1 (NUTS-2 in the past) areas (Martelli,1998; Lima *et al.*, 2009). Each area is treated separately in the sense that the higher variances affecting a given area do not affect the other ones. The influence of the major areas with likely higher variances is thus reduced while that of the less industrialized areas is conversely increased allowing to enlarge the sample size in "weak" strata. To define n_k for k=1,…,4 (i.e. for each of the four NUTS-1 areas for Italy) a simple iterative process is applied. A desired precision d of the total employment estimate corresponding to the given total sample size n=4000 is firstly calculated and this value is then initially used to obtain the sample size for each area. The process is repeated by slightly increasing/lowering d until the sum of the four n_k areas sample sizes approximates the target overall sample size n. In our case this process leads to identify a precision equaling $d \approx 2.2\%$ which, applied to the sample areas, leads to four sub-samples which more properly sum up to n=3951, contributing to the total sample size according to the following area sizes: n_1=34%, n_2=34%, n_3=17%, n_4=15%. This per-stratum geographical distribution further allows for not moving too much away from the real firm population percentages in the selected areas (respectively equaling 35%, 31%, 19%, and 15%. See Table 3.1).

The variability is measured on the stratum's number of persons employed, and a Winsorisation (α=5%) is applied to the standard deviations. For each area, the upper/lower 5% values of σ_{lk} are set equal to the nearest lower/upper value. This technique is usually applied when dealing with partial frames or incomplete populations, where data on variances can be missing. Some sectors (e.g. the transport sector) often present very high variances. As a consequence, more units than necessary might be included in the final sample from these sectors. The stemming sample is often characterized by a higher precision but also presents an over-concentration of sample units in the high-variance strata. The Winsorisation process allows avoiding this undesirable effect by smoothing of extremes values. In this case, being l (l=1,...,L) the generic stratum into area k formed by sector and firm size only, then, according to the Neyman method, the l^{th} stratum size for the k^{th} area (for k = 1, …, K) can be obtained in a similar way to that used to obtain formula [4.3], with H, the total number of strata, being equal to $L \times K$.

The Winsorisation and the optimal strata allocation applied separately for the four geographical areas allow to overcome the drawback of the Neyman method to "over-allocate" units in strata with higher variances, and conversely to "under-allocate" units in areas with weak industrial sectors. These are quite features of the Italian economic system: the over-allocation effect in the North of the country, the under-allocation effect in the Centre and the South.

At the overall sample level, these adjustments show a satisfactory performance, very similar to the classical Neyman design performance (see Chiodini *et al*., 2010), with a satisfactory coverage also for less industrialized areas. This allocation method has successfully been applied for over a decade to the ISAE BCS with a local breakdown at NUTS-2 level with 19 regions considered.

With the aim of reducing the number of strata, the NUTS-1 breakdown (4 partitions) is considered and, above all, a further bound is tested. Practice often suggests that in strata with large firms it is very difficult to achieve a sampling fraction f greater than 20%, while the Neyman approach, in presence of highly volatile strata, often requires census investigations. While the ISAE methodology without any size bound showed satisfactory results, in this paper we wish to test a more realistic sample by including also the $f_h \leq 0.2$ constraint. This bound implies a reduction of the sample size in the most volatile strata with larger firms, given the total sample size n. Using the same technique as above, the total sample size slightly changes, summing up to n=3977.

****

Algorithm

Redefine n_h as in equation [4.3], then use Algorithm 3.

****

4.5 The Power Compromise Allocation

Bankier's (1988) proposal is a compromise between the Neyman allocation and the CV allocation. Letting $CV_h = \sigma_h/\bar{Y}_h$ be the stratum CV, the stratum size for the power allocation (Bankier, 1988) is:

$$n_h^* = n \frac{CV_h X_h^q}{\sum_{h=1}^H CV_h X_h^q} \qquad h=1,\ldots,H. \qquad [4.7]$$

Bankier defines X_h as some measure of size or importance of stratum h and q as a constant in the interval $[0,1]$. The choice of X_h and the values of q taken between 0 and 1 provide a family of compromise allocations. In particular, with q=0 we have the CV allocation, whereas with q=1 and $X_h = N_h \bar{Y}_h = T_h$ we have the Neyman allocation.

In order to better perform the comparison between different allocation methods, we also consider the CV allocation. Furthermore, we consider the values for q proposed by Bankier (1988, p.176), i.e. $q = 0.3$ and $q = 0.5$. With respect to the values for X_h we will consider both $X_h = T_h$ (CV-Neyman family) and $X_h = N_h$, being the latter a weight structure generally used in other methods considered in this paper.

In general, this procedure can be formulated in algorithmic form as follows:

****

Algorithm

Use formula [4.7], define n_h as in equation [4.3], then apply Algorithm 3.

****

4.6 The Standard Deviation Allocation

Using Bankier's proposal with q=1 and $X_h = \bar{Y}_h$ we have the standard deviation (STD) allocation:

$$n_h^* = n \frac{\sigma_h}{\sum_{h=1}^H \sigma_h} \qquad h=1,\ldots,H.$$

In general, this procedure can be formulated in algorithmic form as follows:

****

Algorithm

Like in Bankier's method, define n_h as in equation [4.3], then apply Algorithm 3.
****

4.7 The PAUST Compromise Allocation

Costa's proposal is a compromise between proportional allocation and uniform allocation (here called PAUST – Proportional Allocation with Uniform Stratum Threshold). Letting $W_h = N_h/N$ be the relative size of stratum h for h=1,…,H, the stratum size for the PAUST allocation (Costa et al, 2004) is:

$$n_h = k\,(nW_h) + (1-k)\left(\frac{n}{H}\right) \qquad h=1,\dots,H, \qquad [4.8]$$

where k is a constant in the interval $[0,1]$. The choice of the values for k taken between 0 and 1 provides a family of compromise allocations. In particular, with k=0 we have the uniform allocation, $(n_h = n/H)$, with k=1 we have the proportional allocation, $(n_h = nW_h)$.

Formula [4.8] can also be written as follow:

$$n_h = \frac{n_1}{H} + n_2 W_h$$

where $n_1 = (1-k)\,n$, $n_2 = k\,n$, and obviously $n = n_1 + n_2$. In this way, Costa's formula can be regarded similarly to other methods, highlighting the uniform threshold. For example with $n_1 = 3$ we will write $\text{PAUST}_{(u=3)}$ in the results from the simulation.

Formula [4.8] needs to be modified when $n/H > N_h$ for some h in a set of strata $\mathcal{B}$. The modified stratum allocation is:

$$n_h = k\,(nW_h) + (1-k)\tilde{n}_h, \qquad h=1,\dots,H \qquad [4.9]$$

where:

$$\tilde{n}_h = \begin{cases} \frac{n-\sum_{h\in\mathcal{B}} N_h}{H-m} & h \in \mathcal{A} \\ N_h & h \in \mathcal{B} \end{cases}$$

and m is the number of strata in set $\mathcal{B}$.

If some h are in set $\mathcal{B}$, $n = \sum_h n_h$ is not guaranteed. In this case procedure can be formulated in algorithmic form as follows:

****

Algorithm

Use Algorithm 1 for the uniform allocation part, then Algorithm 2 for the proportional allocation part.

****

4.8 The ROAUST Compromise Allocation

In the context of the domain analysis, an important result has proved to be the Robust Optimum Allocation with Uniform Stratum Threshold (ROAUST; see Chiodini *et al.* 2008), as compared to the classical methods usually applied. In this paper we wish to test its efficiency also in comparison with a new version of the ISAE technique.

Given the desired total sample size $n = \sum_h n_h$, we first apply the uniform allocation by sampling n_1 units ($n_1=\alpha n$ with $\alpha \in [0,1]$) so that for each *h-th* stratum the sample size becomes $n_{1h}=n_1/H$. The Neyman allocation is then applied to the remaining n_2 units, such that $n_2=n-n_1$. Hence, in each h stratum, the sample size is given by:

$$n_h = n_{1h} + n_2 \frac{N_h \sigma_h}{\sum_{h=1}^{H} N_h \sigma_h}. \qquad [4.10]$$

When α=0, $n_{2h}=n_h$ (obtaining in this way the optimal allocation). When α=1, $n_{1h}=n_h$ (obtaining in this way the uniform allocation). Among all the possible values n_1 can assume in each stratum, in this paper two values are proposed, namely n_1=3 and n_1=9. The first is a constraint deriving from the minimum stratum size required to guarantee the privacy of the respondents (whereas with at least three units in a stratum it is not possible to identify the participating firms); with the latter, corresponding to an α value of about 50%, a balanced stratum size is achieved by assigning equal importance both to the stratum information (i.e. a fixed number of units required within each stratum) and the allocation optimality, by allocating the remaining 50% units proportionally to the size and strata heterogeneity. After several empirical simulations, these two values have been shown to be the most efficient for this sample (Chiodini *et al*. 2010). This procedure can be formulated in algorithmic form as follows:

****

Algorithm

Use Algorithm 1 for the uniform allocation part, then Algorithm 3 for the optimal allocation part.

****

4.9 The IPNLP Allocation

A non-linear programming (NLP) method is also considered to obtain an allocation to strata that minimizes the total sample size *n* subject to specified tolerances on the CV of the strata and population mean estimators. The NLP allocation class can be considered essentially a general constrained and non-linear optimized allocation (see Choudhry *et al.*, 2012).

In the simulation study we will use a modified version of the NLP allocation, *Interior Point Non Linear Programming* (IPNLP) solver, which uses a Quasi-Newton Interior Point (IPQN) method. IPNLP-IPQN can efficiently solve medium size optimization problems (see SAS Institute Inc., 2010 for details). For example, it is recommended for problems whose second derivatives are computationally expensive to compute.

Let $\mathbf{n}=(n_1,\ldots,n_H)$. With IPNLP we aim at minimizing the total sample size:

$$g(\mathbf{n}) = \sum_{h=1}^{H} n_h. \qquad [4.11]$$

Similarly to Choudhry *et al.*, we use the root of the relative variances, the CV of $\bar{y}_h$:

$$CV(\bar{y}_h) = CV_h\sqrt{\frac{N_h - n_h}{n_h N_h}} \qquad h=1,\ldots,H, \qquad [4.12]$$

and the CV of $\bar{y}_{st}$, the estimated population mean:

$$CV(\bar{y}_{st}) = \sqrt{\sum_{h=1}^{H}\frac{N_h - n_h}{n_h N_h} W_h^2 \frac{\sigma_h^2}{\bar{Y}^2}}, \qquad [4.13]$$

where $W_h = N_h/N$ is the relative size of stratum h ($h=1,\ldots,H$).

We use SAS IPNLP (Interior Point Non Linear Programming) solver with the IPQN option to find the optimal n_h that would minimize [4.11] subject to:

$$CV(\bar{y}_h) \leq CV_{0h} \qquad h=1,\dots,H \qquad [4.14]$$

$$CV(\bar{y}_{st}) \leq CV_0 \qquad [4.15]$$

$$1 \leq n_h \leq N_h \qquad h=1,\dots,H \qquad [4.16]$$

where CV_{0h} and CV_0 are specific tolerances on the CV for the stratum sample mean $\bar{y}_h$ and the estimated population mean $\bar{y}_{st}$ respectively. Note that the constrain $n_h \geq 1$ implies a uniform stratum threshold equal to 1.

For the ease of comparison, given *n*=3981, and assumed the CV_0 equal to the RRMSE of the ROAUST$_{(u=9)}$:

$$CV_0 = {}^{ROAUST_{(u=9)}}RRMSE \quad = 0.69\%$$

we fix the correspondent value for CV_{0h}:

CV_{0h}= 11.60%.

4.10 The Neyman-Bethel Multivariate Allocation

A multivariate allocation on qualitative data is also considered, as the BCS investigates various phenomena originating also qualitative data (see also Kozak, 2006).

This allocation method represents an extension of the multivariate case of the Neyman allocation, as the number of variables involved is *j*>1. It is performed according to Bethel's (1989) formulation and uses the Chromy (1987) iterative algorithm to solve it (Falorsi *et al.*, 1998; ISTAT, 2004). In a recent research (Lima *et al.*, 2009) the multivariate allocation was applied using variances of quantitative variables, namely the number of persons employed and the turnover, but no appreciable gain in efficiency was achieved with respect to the classical Neyman univariate approach.

In this research we test a further approach: the selected variances are those of the three qualitative key variables forming the industrial composite confidence indicator (namely: order assessments, production expectations and stock level). It is worth recalling that in this case the selected variables can be only categorical, not allowing proper comparisons with the other considered allocation methods. In particular, given that in our exercise we are dealing with qualitative variables (with three categories), the stemming strata variances tend to become uniform, so that their contribution in strata definition would be limited, likely leading to results more similar to those of the proportional allocation by construction.

4.11 Some remarks on the selected methods

It is important to recall here a fundamental property of the Neyman's allocation, which arises when comparing the variances (V) of the estimated means obtained from different sampling designs, like the simple random sampling, the stratified random sampling with proportional and optimal allocation. In fact, it is possible to show that:

$$V_{optimal}(\bar{y}) \leq V_{proportional}(\bar{y}) \leq V_{random}(\bar{y}), \quad [4.17]$$

having the optimum allocation estimates minimum variance (Cochran, 1977 sec. 5.6). Here we focus on stratified allocations (which are more efficient), also introducing the uniform, compromise and IPNLP allocation.

As the Neyman allocation, also the uniform allocation is disproportionate, and if the strata are also domains, as in our case, equal n_h's allow for an improvement in the estimates (Kish, 1965, section 3.6). We wish to test whether and to what extent the combination of disproportionate allocations as well as the introduction of strong bounds could give better results.

In our analysis we exclude allocation methods using take-all-strata methods. Generally, in take-all-strata methods all the elements of the population are included in the sample (Hidiroglou, 1986; Lavallée and Hidiroglou, 1987, Baillargeon and Rivest 2009). This is often the case where a few strata comprise large firms. Even though widely applied in business surveys to reduce the variability in highly skewed strata, the take-all strata approach has never been explicitly applied to the Italian BCS up to 2010. Motivations behind this choice can be ascribed to the following circumstances: i) the survey has been conducted for a long time on a voluntary basis, as data collection from freely-willing-to-answer firms has been preferred in order to privilege the quality of the information of the collected questionnaires; ii) past experience has shown that only about 20% of large firms voluntarily agrees to take part to the survey.

Reasons for nonresponse are essentially the following: i) ISAE had not enough authority to solicit the respondents; ii) several persons from different departments are usually involved in the process of filling in the different parts of the questionnaire; iii) the presence of statistical attritions, as firms gradually become tired of answering a monthly timing survey; iv) the monthly timing of the survey which prevents from systematic reminders to non-responding firms and from reaching a higher response rate; v) some topics may play a strategic role and are not voluntary disseminated. Finally the mandatory nature of the survey participation may lead to a hasty filling of the questionnaire, reducing its quality.

It is also worth highlighting here that the use of a quantitative auxiliary variable, while dealing with a survey mainly aimed at collecting several categorical opinions, is bound to the features of BCS surveys. As pointed in sec. 2 the importance of the answers is assumed to depend on the size of the investigated firms (UNSD, 2013). The following considerations may help to explain the process: i) first of all, confidence represents a timely proxy of quantitative variables like the GDP dynamic. In fact, in the case of business confidence surveys the validation exercise for the confidence series is usually performed by comparing these series with the corresponding quantitative series (once suitable treated: i.e. seasonally adjusted and once the cyclical component has been extracted). Furthermore, from a more statistical point of view: ii) the importance of the collected categorical results (within the three-level Likert-scale for each question) is roughly quantified in the subsequent weighting process by the firm size in terms of persons employed, from a register source in order to allow comparisons between different designs (the larger the firm the more the answers affect the final outcomes; see OECD, 2003, sec 114 p 36); iii) this weighting system allows to strengthen the adopted sample design as the BCS survey is carried out also for collecting quantitative data on investments and for checking several up-to-date phenomena (e.g. labour market, credit restrictions, etc.).

On the other hand, the use of the number of persons employed as auxiliary variable allows some quantitative estimation and using register data allows comparisons between different sampling designs.

5. Simulation

The objective of this section is to illustrate the framework used to compare different allocation strategies via Monte Carlo simulation. Monte Carlo simulation allows both to ignore data distribution assumptions and obtain a thorough comparison of the allocation methods in terms of empirical performance.

Our simulation is structured with the aim of separating the selection, the allocation and the inferential process. This can be achieved via a *Monte Carlo-Sequential-Selection-Allocation* (MC-SSA) process (Chiodini *et al*. 2010) which can be linked to the Permanent Random Number (PRN) technique proposed by Ohlsson (1995). PRN allows to optimize the process of selection of the units in repeated sampling from the same frame, maximizing the proportion of overlapping units between the compared allocation methods.

Let P be a population subdivided in H strata, $N_1,\ldots,N_h$ the population strata sizes, $n_1,\ldots,n_h$ the sample strata sizes, with $N = \sum N_h = 86447$ being the population size, and $n = \sum n_h \cong 4000$ the target overall sample size. The MC-SSA process works as follows:

1. replicate 1000 times the original population subdivided in H strata. Call these replicates $P_1,\ldots,P_r,\ldots,P_{1000}$;
2. for each P_r:
    a. assign distinct PRNs to the units of the population;
    b. sort the units of each stratum in ascending order with respect to the assigned PRNs;
    c. select the sample units in each population stratum according to each allocation method to be compared. Sample sizes from each allocation method are almost the same (i.e. $n \cong 4000$). By doing this, in each strata and across all the considered allocation methods, the maximum number of overlapping units is guaranteed.

d. obtain the sample estimate $\bar{y}_r$ of the mean number of persons employed;

e. obtain 226 non-empty mean stratum estimates $\bar{y}_{rh}$;

3. from the 1000 estimates of $\bar{y}_r$ obtain the RRMSE;
4. from the 226 strata, obtain the 226 strata RRMSEs.

With this simulation plan we guarantee that:

- using Ohlsson's approach, the differences resulting in the estimates are scarcely influenced by the non-overlapping units across different allocation methods;
- differences in the quality of the estimates are scarcely influenced by the different sample sizes, because they differ slightly when implementing the allocation methods.

These two characteristics of our simulations allow highlighting only the effects due to the sample allocation design.

6. Results

The morphology of the Italian industrial distribution is characterized by a large fragmentation due to the massive presence of a large amount of small industrial firms with a limited number of persons employed. The distribution of industrial firms with regard to their size is highly skewed, and even if the survey discards companies with fewer than 10 persons employed, it can be observed that the truncated distribution obtained with the logarithm of the number of persons employed per company still shows a high degree of positive asymmetry and the resulting distribution is markedly leptokurtic (Figure 6.1).

MC simulations are needed to verify the properties of the estimators derived from the ROAUST method in comparison with the estimators from other methods proposed in the literature, or usually adopted in business surveys, as analytical check seems not easy to obtain in this case.

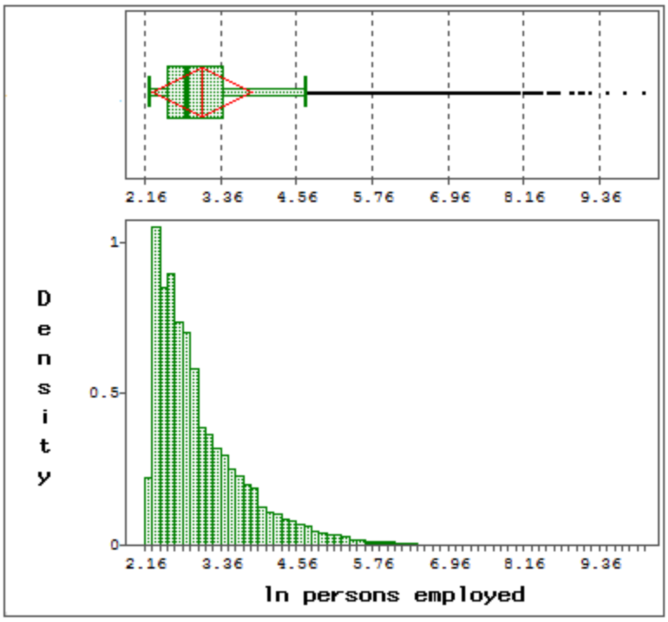

Figure 6.1 Distribution of firms by the natural log of persons employed

Source: Data processing on ISTAT 2007 data.

6.1 Overall population results

In order to compare the allocation methods, we use data on the persons employed as reported by the Asia frame, and, as a measure for comparison, the well-known, relative root mean square error (RRMSE) defined by:

$$\text{RRMSE} = \sqrt{\left(\frac{Bias}{\bar{Y}_r}\right)^2 + CV_r^2}, \qquad [6.1]$$

where *Bias* is estimated as $\bar{Y} - \bar{Y}_r$, $\bar{Y}$ is the population mean, $\bar{Y}_r$ is the empirical mean obtained over all the replicates according to the MC simulation process, $CV_r = \sigma_r / \bar{Y}_r$, with σ_r being the empirical standard error obtained across the replicates.

The *RRMSE* is lower in the Neyman/ISAE allocation when applied to the overall population (see Table 6.1). This is a well-known result under normality and is still valid under non-normality since it takes into account the stratum heterogeneity and the stratum size at the same time, whereas some of the other methods do not retain them. The firms' distribution exhibits a high degree of positive asymmetry and the inevitable presence of influential observations in the economic system. They are representative of the characteristics of the population and therefore not negligible in terms of selection of units. The Neyman approach takes into account the influential observations (Lee, 1995). The inclusion/exclusion of these extreme values affects the estimates.

Figure 6.2 provides a graphical representation of the accuracy of the different methods considered via MC simulation. It can be easily noted that the proportional method produces a positively asymmetric MC distribution. Table 6.1 reports for the overall population the sample allocation $n \cong 4000$, the associated relative Bias ($|Bias/\bar{Y}_r|$) and the relative root mean square error for different allocation methods: the benchmark allocations (i.e. the modified Uniform, Proportional and Neyman allocation methods), the compromise allocation (i.e. Bankier, PAUST and ROAUST allocation methods) and the IPNLP allocation method. First of all, starting from $n \cong 4000$, we find n=3984 for the Uniform allocation and the proportional allocation. For the other allocation methods, with the exception of the Bethel and the ISAE allocation methods which are considered apart, we fix $n \leq 3984$. In particular n=3983 for the STD, Neyman and PAUST allocation methods; n=3982 for the Bankier method; n=3981 for the ROAUST and IPNLP methods. It shows that the IPNLP respect a specified CV_0 constraint that was fixed considering the $\text{ROAUST}_{(u=9)}$. Therefore, let n=3981 and $CV_0 = 0.69\%$; then $CV_{0h} = 11.60\%$.

From Table 6.1 and Figure 6.2 it is clear that the modified proportional allocation and the Costa's proposal (i.e. $\text{PAUST}_{(u=9)}$ and $\text{PAUST}_{(u=3)}$) are not suitable in terms of RRMSE, since its RRMSE is greater (more than three times) than that of the Neyman allocation method. We also observe that the choice of the Bankier (q=0.5; X_h=N_h), the modified STD and the Uniform allocation imply values for the RRMSE which almost double those of the Neyman allocation. Finally, most of Bankier's RRMSEs are significantly larger than those of the Neyman allocation method.

Considering allocation methods with an overall population RRMSE less or equal to 0.70% we have five interesting proposal: the modified Neyman method with RRMSE = 0.63%; the $\text{ROAUST}_{(u=3)}$ method with RRMSE = 0.64%, $\text{ROAUST}_{(u=9)}$ and the IPNLP method with RRMSE = 0.69%, the Bankier(q=.5; X_h=T_h) method with RRMSE = 0.70%.

Table 6.1
MC simulation results: Relative errors (overall population, 1000 replicates)

	Sample size	$\lvert Bias / \bar{Y}_r \rvert$	*Overall RRMSEs*
IPNLP$_{(U=1)}$	**3981**	**0.0001**	**0.0069**
STD = BANKIER (q=1; $X_h = \bar{Y}_h$)	3983	0.0008	0.0144
ISAE BOUNDED	3977	0.0006	0.0250
NEYMAN - BETHEL	3993	0.0004	0.0264
NEYMAN / ISAE * = ROAUST$_{(U=0)}$ = BANKIER (q=1; X_h=T_h)	**3983**	**0.0000**	**0.0063**
ROAUST$_{(U=3)}$	**3981**	**0.0000**	**0.0064**
ROAUST$_{(U=9)}$	**3981**	**0.0000**	**0.0069**
UNIFORM = ROAUST$_{(U=21)}$ = PAUST$_{(U=21)}$	3984	0.0003	0.0179
PAUST$_{(U=9)}$	3983	0.0006	0.0248
PAUST$_{(U=3)}$	3983	0.0001	0.0355
PROPORTIONAL$_{(U=1)}$ = PAUST$_{(U=1)}$	3984	0.0007	0.0520
BANKIER (q=.5; X_h=N_h)	3982	0.0000	0.0130
BANKIER (q=.3; X_h=N_h)	3982	0.0000	0.0099
BANKIER (q=.0; X_h=N_h=T_h) = CV	3982	0.0001	0.0098
BANKIER (q=.3; X_h=T_h)	3982	0.0001	0.0079
BANKIER (q=.5; X_h=T_h)	**3982**	**0.0001**	**0.0070**

Source: Data processing on ISTAT, 2007 data.
Note: (*) Results for ISAE and Neyman allocations are very similar and hence are gathered in a single column.

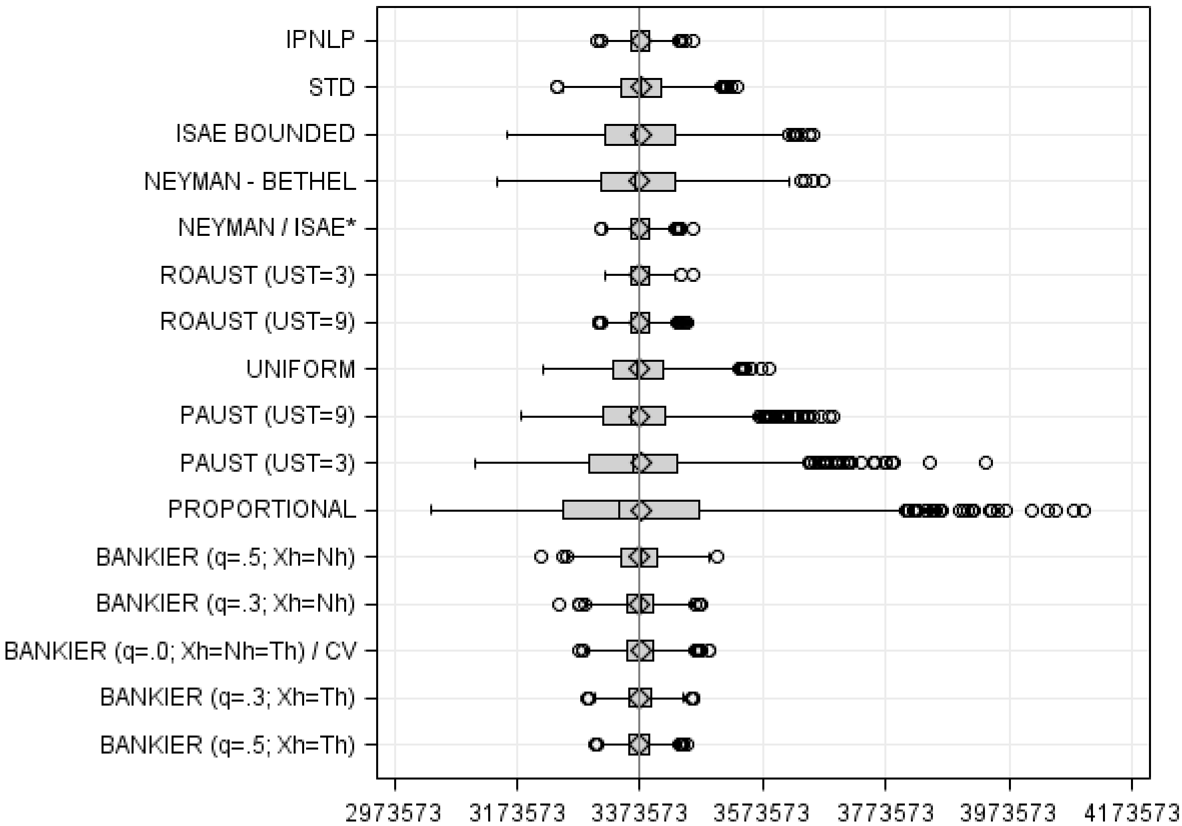

Figure 6.2 MC boxplots: total (1000 replicates)
Source: Data processing on ISTAT 2007 data. Y-axis: Number of persons employed.
Note: In Figure 6.2 and 6.3, the vertical lines in the diamonds mark the means (Tukey, 1977).

6.2 Stratum results

In order to consider the means μ_h – that refer to different strata – we apply formula [6.1] that gives the RRMSE. It can be observed (Table 6.2 and Figure 6.3) that the $ROAUST_{(u=9)}$ method, although with a little loss in terms of precision of the overall population estimator (i.e. $ROAUST_{(u=9)}$ RRMSE = 0.0069 versus Neyman RRMSE = 0.0063), results to have the higher accuracy within the strata (i.e. the maximum strata $ROAUST_{(u=9)}$ RRMSE is equal to 0.1588 versus the maximum strata Neyman RRMSE, that is 0.5809).

Table 6.2 reports the number of missing strata for each method, the maximum of the stratum relative |Biases| and the maximum of the stratum RRMSEs. Considering allocation methods with an overall population RRMSE lower or equal to 0.70% (i.e. the modified Neyman, $ROAUST_{(u=3)}$, $ROAUST_{(u=9)}$, IPNLP and Bankier($q=.5;\ X_h=T_h$)), and, considering the maximum of the stratum RRMSE for $ROAUST_{(u=9)}$ as the reference value (base=100), as expected we observe an index more than three times larger (index=366) for the Neyman allocation, an index half time larger (index=157) for $ROAUST_{(u=3)}$ and Bankier($q=.5;\ X_h=T_h$), and finally an unexpected index a quarter larger for IPNLP (index=125).

Figure 6.3 reports the strata RRMSEs for each method. We can observe that for the IPNLP and the $ROAUST_{(u=9)}$ methods there are slight differences: the error for IPNLP is centered and concentrated on 12% whereas the large part of the distribution of $ROAUST_{(u=9)}$ stratum errors is concentrated below 12%. Finally, even if $ROAUST_{(u=9)}$ is better than IPNLP in 49% of the strata, it is equal in 23% of the strata and worse in only 28% of the strata. Also when considering the distribution of the stratum RRMSE quartiles (Table 6.3), $ROAUST_{(u=9)}$ results always equal or lower than IPNLP. The reason for this can be seen when comparing the frequency distributions of the two methods (Figure 6.4). For example:

- Mode of $ROAUST_{(u=9)}$ RRMSE = [0.00-0.01] vs Mode of IPNLP RRMSE = [0.11-0.12];
- Max of $ROAUST_{(u=9)}$ RRMSE = [0.15-0.16] vs Max of IPNLP RRMSE = [0.19-0.20].

Table 6.2
MC simulation results: Relative errors among strata (1000 replicates)

	Missing strata #	*Max of strata relative |Biases|*	*Max of strata RRMSEs*	*Overall RRMSEs*
IPNLP$_{(U=1)}$	**0**	**0.0109**	**0.1978**	**0.0069**
STD = BANKIER (q=1; $X_h=\overline{Y}_h$)	4	0.0167	0.2181	0.0144
ISAE BOUNDED	6	0.0413	1.2881	0.0250
NEYMAN - BETHEL	0	0.0352	1.0115	0.0264
NEYMAN / ISAE * = ROAUST$_{(U=0)}$ = BANKIER (q=1; X_h=T_h)	**11**	**0.0155**	**0.5809**	**0.0063**
ROAUST$_{(U=3)}$	**0**	**0.0155**	**0.2495**	**0.0064**
ROAUST$_{(U=9)}$	**0**	**0.0115**	**0.1588**	**0.0069**
UNIFORM = ROAUST$_{(U=21)}$ = PAUST$_{(U=21)}$	0	0.0112	0.3966	0.0179
PAUST$_{(U=9)}$	0	0.0292	0.6743	0.0248
PAUST$_{(U=3)}$	0	0.0362	1.0105	0.0355
PROPORTIONAL$_{(U=1)}$ = PAUST$_{(U=1)}$	0	0.0550	1.6587	0.0520
BANKIER (q=.5; X_h=N_h)	5	0.0155	0.3184	0.0130
BANKIER (q=.3; X_h=N_h)	5	0.0112	0.1520	0.0099
BANKIER (q=.0; X_h=N_h=T_h) = CV	5	0.0103	0.1187	0.0098
BANKIER (q=.3; X_h=T_h)	5	0.0103	0.1526	0.0079
BANKIER (q=.5; X_h=T_h)	**5**	**0.0116**	**0.2495**	**0.0070**

Source: Data processing on ISTAT, 2007 data.
Note: (*) Results for ISAE and Neyman allocations are very similar and hence are gathered in a single column.

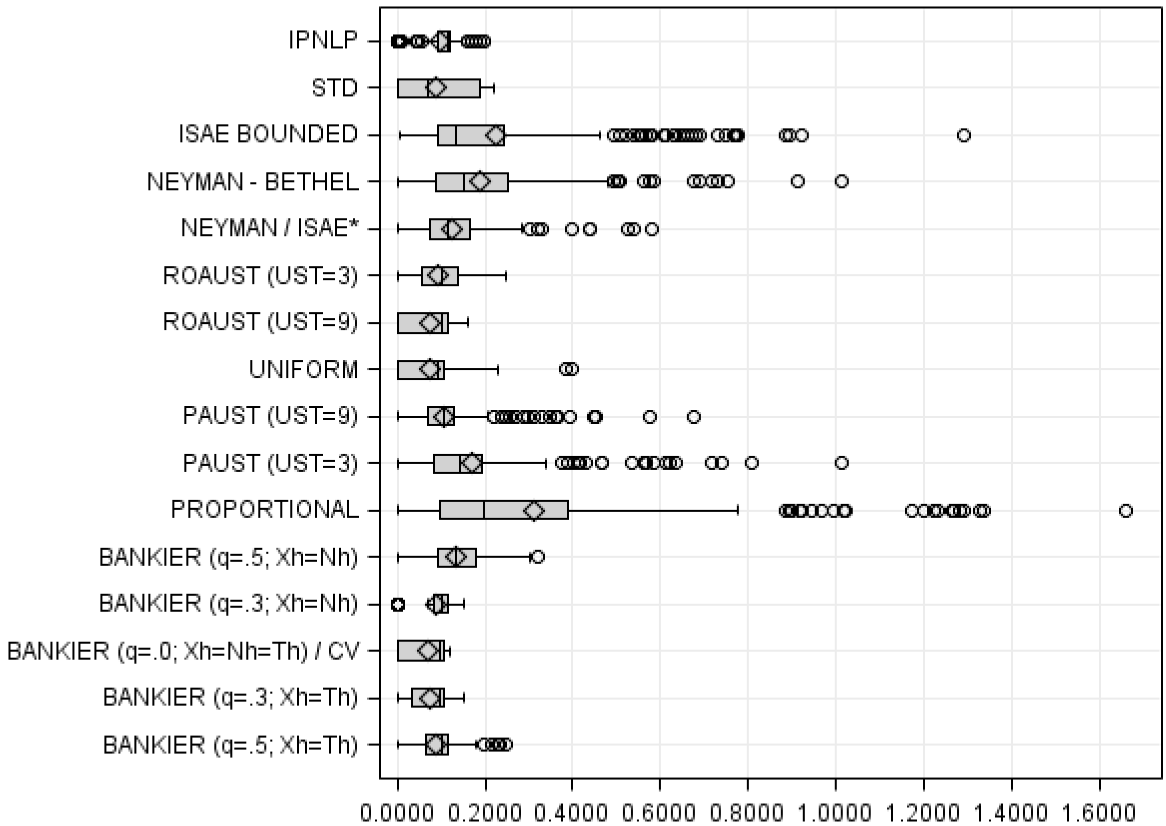

Figure 6.3 RRMSE by strata (# 228, where 226 strata are populated, 1000 replicates)
Source: Data processing on ISTAT, 2007 data.

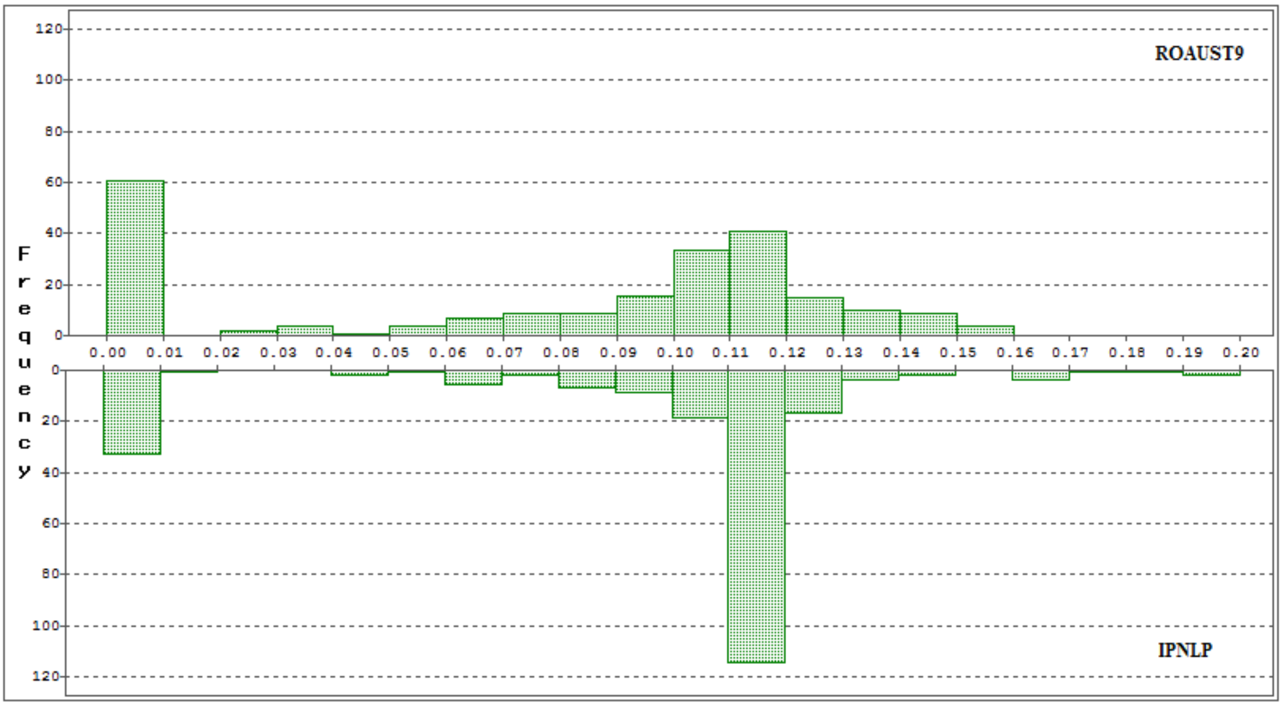

Figure 6.4 RRMSE by strata, IPNLP vs ROAUST$_{(u=9)}$ (# 228, where 226 strata are populated, 1000 replicates)
Source: Data processing on ISTAT 2007 data.

Table 6.3 RRMSE by strata, IPNLP vs ROAUST$_{(u=9)}$ (# 228, where 226 strata are populated, 1000 replicates)

RRMSE	Strata %	Persons employed %
IPNLP-ROAUST9 > 0	**49%**	**38%**
IPNLP-ROAUST9 = 0	23%	24%
IPNLP-ROAUST9 < 0	28%	**38%**
Total	100%	100%

RRMSE	IPNLP	ROAUST$_{(u=9)}$
Max	0.1978	0.1588
Q3	0.1178	0.1150
Med	0.1144	0.0996
Q1	0.0931	0.0000
Min	0.0000	0.0000
Range	0.1978	0.1588

Source: Data processing on ISTAT, 2007 data.

Finally, given the difficulty in reaching all sample units in the small strata (i.e. to apply the so called "take all" strata technique, where $n_h = N_h$ for some strata), we also tested a very strong operational constraint applied to the ISAE approach: that is, we assume a very sharp upper bound (e.g. 20% of the stratum units for the large sized strata. The results, however, suggest that constraints may be introduced only up to a certain extent. Chiodini *et al*. (2010) proved that, by introducing a sampling fraction not higher than 50% (together with a lower bound of 3 units per stratum), the strata accuracy of the ROAUST methods is not lowered.

However, in the present ISAE case the constraint seems to be too tight as the results, even always better than for proportional allocation, are not satisfactory.

7. Discussion

The sense of the interdisciplinary approach used in this paper can be well summarized by the following Martini's quote (2004, p. 10): "*Statistics cannot be used in reality if statisticians are not at the same time economists and sociologists and, above all, are not driven by a problem-solving approach which must start from the entities forming the socioeconomic populations in their work*". Sector surveys often require solutions that are not extensively available in the literature. In addition to the usual sample bias connected to the sample lists and the information retrieved via questionnaires, one has to deal with the estimation of population variables that are not normally distributed.

Our proposal comes from the observation of reality with a pragmatic approach, and leads to methodological developments to overcome problems like those of imperfect frames and heterogeneity in strata. The ROAUST class can be considered a Neyman's domain allocation, since it allows both for optimal allocation and stratum information (i.e. by construction it is necessary to fix a number of units required within each stratum).

The need of updating the ISTAT BCS survey sample allocation to the NACE Rev. 2 classification gave us a twofold opportunity to propose a compromise allocation method, and to compare it with other popular allocation designs. To accomplish this aim, an *ad hoc* simulation tool (i.e. MC-SSA) was set up and applied.

However before we resume the main characteristics of the chosen allocation methods arisen from our simulation study, we observe that the use of employment as auxiliary variable is conditioned on the administrative archive used.

In summary, this paper points out that some criticalities affect the goodness of the Neyman's allocation method:

1. by construction, the Neyman/ISAE allocation can not allocate units in some strata
2. the Neyman multivariate allocation according to Bethel is not suitable for BCS sampling definition.

Furthermore, two points about benchmark allocation methods are worth highlighting:

1. the proportional allocation method, although widely used by sectorial operators, is not suitable when firm size are used as stratum variable, as usually applied in european business surveys;
2. the uniform allocation performs better than proportional allocation both at domain and overall level.

The latter result is interesting also for the simplicity of this allocation method. The efficiency of the uniform allocation is better than that of the proportional allocation when some conditions are met, for example when a decrease in stratum size corresponds to an increase in stratum variability. This is confirmed by our results in which the correlation between the stratum size and stratum variability is equal to -0.2. On the contrary, Choudry *et al*. found that the proportional allocation has a better performance than the uniform allocation. In our view this is also due to a positive correlation (equal to +0.8) between the stratum size and the stratum variability.

Moreover, our simulation results highlight some further points:

1. the PAUST compromise allocation (Costa *et al*., 2004) performs similarly to the proportional allocation;
2. the Bankier compromise allocation (Bankier, 1988) performs well in many cases;
3. the best results in terms of overall and stratum efficiency are those of ROAUST method. The simultaneous presence of these two features makes the ROAUST optimal for domain analyses (e.g. regional and sectorial).
4. the Nonlinear Programming (NLP) method (Chaudry *et al*., 2012) - in the more flexible version of the IPNLP here considered – minimizes the total sample size, subject to a specified tolerance on the CVs. However, it can be noted that the MC-SSA simulation for the IPNLP method has highlighted a range of the stratum RRMSE larger than that of the ROAUST method.

The negative correlation between stratum size and stratum variability is a common characteristic of the European business surveys, since the firm size is usually requested as a stratification variable. All in all, thanks to its functional simplicity deriving from its additive form that allows to exploit the advantages of both the uniform and the Neyman components, the ROAUST is naturally candidate method for business surveys.

In countries where business registers are not reliable the ROAUST method is even more useful since the stratum variability cannot be known *ex-ante*. In fact, the uniform component can be used to estimate the population parameters for the Neyman allocation.

References

Baillargeon, S., and Rivest, L. P. (2009). A General Algorithm for Univariate Stratification. *International Statistical Review*, 77(3), 331-344.

Bankier, M. D. (1988). Power Allocations: Determining Sample Sizes for Subnational Areas. *The American Statistician*, 42(3), 174-177.

Bethel, J. (1989). Sample Allocation in Multivariate Survey. *Survey Methodology*, 15(1), 47-57.

Buglielli, T., De Vitiis, C., Barcaroli, G. (2011). MAUSS-R: Multivariate Allocation of Units in Sample Surveys. User and Methodological Manual. ISTAT.

Chiodini, P.M., Facchinetti, S., Manzi, G., Nai Ruscone, M., Verrecchia, F. (2008a). *Metodi e applicazioni per l'inferenza da popolazione finita: imprese e campionamento stratificato a selezione ordinata*. (ESeC Working Paper: ESeC_WP004P_V20080714).

Chiodini, P.M., Manzi, G., Verrecchia, F. (2008b). *Allocazione ottimale robusta con soglia uniforme di strato*. (ESeC Working Paper: ESeC_WP005P_V20080912).

Chiodini, P.M., Manzi, G., Martelli, B.M., Verrecchia, F. (2010a). *Between Theoretical and Applied Approach: Which Compromise for Unit Allocation in Business Surveys*. (paper presented 45th Scientific Meeting of the Italian Statistical Society, University of Padua, Italy).

Chiodini, P.M., Lima, R., Manzi, G., Martelli, B.M, Verrecchia, F. (2010b). Criticalities in Applying the Neyman's Optimality in Business Surveys: a Comparison of Selected Allocation Methods. In: J.Wywiał and W. Gamrot (Eds.), *Survey Sampling Methods in Economic and Social Research* (p. 42-77). Katowice University of Economics Publishing Office, Poland.

Chiodini, P.M., Manzi, G., Martelli, B.M., Verrecchia, F. (2011). *On Computational Aspects of Simulation Methods in the Sample Allocation Framework*. (paper presented at 4th Conference of the European Survey Research Association, Lausanne, Switzerland).

Choudhry, G. H., Rao, J. N. K., & Hidiroglou, M. A. (2012). On Sample Allocation for Efficient Domain Estimation. *Survey Methodology*, 38(1), 23-29.

Chromy, J. (1987). *Design Optimization with Multiple Objectives*. (Proceedings of the Section on Survey Research Methods Section of the American Statistical Association, p.194-199).

Cochran, W.G. (1977). *Sampling Techniques*. New York: John Wiley & Sons.

Costa, A., Satorra, A., Ventura, E. (2004). Using Composite Estimators to Improve both Domain and Total Area Estimation. SORT, 28, 69-86.

Elisson, H, and Elvers, E (2001). Cut-off sampling and estimation, Statistics Canada International Symposium Series – Proceedings. Available at http://www.statcan.ca/english/freepub/11-522-XIE/2001001/session10/s10a.pdf

European Commission (1961). Decision of EC, 15 November 1961. Bruxelles.

European Commission (1997). Commission Decision C(97)2241 of 15 July 1997. Bruxelles.

European Commission (2003). Commission Recommendation of 6 May 2003 concerning the definition of micro, small and medium-sized enterprises, (notified under document number C(2003) 1422, (2003/361/EC). *Official Journal of the European Union*. Bruxelles.

European Commission (2006a). The Joint Harmonised EU Programme of Business and Consumer Surveys. *European Economy*, Special Report No. 5, Bruxelles.

European Commission (2006b). NACE (Nomenclature générale des Activités économiques dans les Communautés Européennes) Rev.2 Classification. *Official Journal* 20 Dec.2006 and *EC Regulation* n.1893/2006 *of the European Parliament and of European Council* 20/12/2006.

European Commission (2006c). Commission Communication COM (2006) 379 of 12 July 2006. Bruxelles.

European Commission, Directorate-General For Economic And Financial Affairs, (2014). *The Joint Harmonised EU Programme of Business and Consumer Surveys User Guide*. Bruxelles. Available : http://ec.europa.eu/economy_finance/db_indicators/surveys/documents/bcs_user_guide_en.pdf [November 2014]

European Union (2003). REGULATION (EC) No 1059/2003 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 26 May 2003 on the establishment of a common classification of territorial units for statistics (NUTS), *Official Journal of the European Union.*

Eurostat (2008). NACE Rev. 2: Statistical classification of economic activities in the European Community, *Methodologies and Working papers*, Luxembourg.

Falorsi, P.D., Ballin, M., Scepi, G., De Vitiis, C. (1998). Principi e metodi del software generalizzato per la definizione del disegno di campionamento nelle indagini sulle imprese condotte dall'Istat, *Statistica Applicata 10(2)*, 235-257.

Hidiroglou, M.A. (1986). The construction of a self-representing stratum of large units in survey design, *The American Statistician 40*, 27-31.

ISCO (Istituto per lo Studio della Congiuntura) (1961). Progetto per un'inchiesta rapida mensile fra i sei paesi della Comunità Economica Europea, *Congiuntura Italiana 12 ISCO.*

ISTAT - National Institute of Statistics (1989). Manuali di tecniche d'indagine, several volumes.

ISTAT - National Institute of Statistics (2004). Software MAUSS (Multivariate Allocation of Units in Sampling Surveys) freely available on-line both in SAS and R version. Available: http://www.istat.it/strumenti/metodi/software/campione/mauss/

ISTAT - National Institute of Statistics (several yearly issues 2005-2011a). Demografia d'impresa, *Statistiche Report.*

ISTAT - National Institute of Statistics (several yearly issues 2005-2011b). Struttura e dimensione delle imprese. Archivio statistico delle imprese attive (ASIA), *Statistiche Report.*

ISTAT - National Institute of Statistics (2005). Censimento dell'Industria e dei servizi. Censimento 2001. Available: http://www3.istat.it/salastampa/comunicati/non_calendario/20050128_01/ [November 2014]

ISTAT - National Institute of Statistics (2007) Struttura e dimensione delle imprese. Archivio statistico delle imprese attive (ASIA), *Database*.

ISTAT - National Institute of Statistics (2009). Classificazione delle Attività Economiche: ATECO 2007. *Metodi e Norme 40*. Available : http://www.istat.it/strumenti/definizioni/ateco/ [November 2014]

James, G., Pont, M. and Sova, M. (2005). *Aspects of Sample Allocation in Business Surveys*. Office for National Statistics, Newport, UK.

Katona, G. (1975). *Psychological Economics.* Elsevier Scientific Publishing Company, New York.

Kish, L. (1965). *Survey Sampling*. Wiley Classics Library, New York.

Kozak, M. (2006). On Sample Allocation in Multivariate Surveys. *Communication in Statistics-Simulation and Computation*, *35*, 901-910.

Kozak, M., Verma, M.R., Zieliński, A. (2007). Modern Approach to Optimum Stratification: Review and Perspectives. *Statistics in Transition - new series*, *8(2)*, 223-250.

Kozak, M., Jankowski, P. (2008). Allocation Constraints in Stratification. *Communication in Statistics-Simulation and Computation*, *37*, 1763-1775.

Lavallée, P. and Hidiroglou, M.A. (1987). On the stratification of skewed populations. *Survey Methodology*, *14*, 33-43.

Lee, H. (1995). Outliers in Business Surveys. In, B.G. Cox, D.A. Binder, B.N., Chinnappa, A., Christianson, M.J., Colledge, P.S., Kott, (Eds.), *Business Surveys Methods* (p. 503-526). New York: Wiley.

Lima, R., Martelli, B.M, Rocchetti, G. (2009). *The ISAE Business Tendency Surveys: Upgrading the Manufacturing Survey Sample Design*. (Papers presented at ITACOSM09 (First Italian Conference on Survey Methodology - Book of Abstract p.247- 250), Siena, Italy.

Malgarini, M., Margani, P., Martelli, B.M. (2005). New design of the ISAE Manufacturing Survey. *Journal of Business Cycle Measurements and Analysis*, *1*, 125-142.

Martelli, B.M. (1998). Le inchieste congiunturali dell'ISCO: aspetti metodologici". In: *Le inchieste dell'ISCO come strumento di analisi della congiuntura economica. Rassegna di lavori dell'ISCO*, XV,3, chap.1. Available : http://ideas.repec.org/p/pra/mprapa/16331.html [November 2014]

Martini, M. (2004). La Statistica. In: M. M. Pelagatti, (Ed.), *Studi in ricordo di Marco Martini*. Giuffrè, Milano.

Neyman, J. (1934). On the Two Different Aspects of the Representative Method: The Method of Stratified Sampling and the Method of Purposive Selection. *Journal of the Royal Statistical Society*, 97, 558–625.

OECD (2003). *Business Tendency Survey: A Handbook*. Paris.

Ohlsson, E. (1995) Co-ordination of samples using permanent random numbers. In Business Survey Methods (eds B. Cox, D. Binder, B. Chinnappa, A. Christianson, M. Colledge and P. Kott), pp. 153-169. New York: Wiley.

Pinca, F. (1990). La Regionalizzazione delle Indagini Congiunturali. In M. Strassoldo (Ed.), *L'Analisi della Congiuntura Economica Locale: Modelli, Metodi e Basi informative*. Padua, CEDAM.

SAS Institute Inc. (2010). SAS/OR® 9.22 User' s Guide: Mathematical Programming. Cary, NC: SAS Institute Inc.

Särndal, C.E., Swensson, B., Wretman, J. (2003). *Model Assisted Survey Sampling*. Springer.

Smith, P., Pont, M., Jones, T. (2003). Developments in business survey methodology in the Office for National Statistics. *The Statistician*, 52, 257-295.

Tukey, J.W. (1977). *Exploratory Data Analysis*, Reading, MA: Addison-Wesley.

UNSD (United Nation Statistical Directorate) (2013). *Handbook on Economic Tendency Survey*. Draft.

Zarnowitz, V. (1992). *Business Cycle, Theory, History, Indicators, and Forecasting*. The University Chicago Press, Chicago.

A1. Appendix: Stratum Statistics

Table A1.1

$\bar{Y}$ by stratum (NUTS-1, firm size, economic sectors), Italy, 2007

NUTS1 and firm size / NACE Sectors	North-West			North-East			Centre			South		
	10-49	50-249	250+	10-49	50-249	250+	10-49	50-249	250+	10-49	50-249	250+
10-12.	18.8	100.6	714.5	18.7	105.7	877.6	17.4	97.7	553.4	17.7	91.0	450.3
13.	20.9	98.6	465.3	18.7	93.9	593.8	17.9	80.6	412.3	18.7	97.4	313.5
14.	17.8	96.3	804.3	17.7	92.5	484.0	18.3	87.9	428.6	18.6	81.8	635.3
15.	17.4	88.1	331.0	19.5	98.4	376.9	18.2	78.6	786.1	18.1	86.5	337.0
16-17.	18.2	96.0	689.4	18.4	97.9	491.4	18.1	88.6	513.5	17.2	93.5	553.0
18.	18.0	94.1	424.3	18.1	96.3	541.8	17.1	91.2	2308.0	16.9	68.8	297.0
19.	22.2	133.6	646.3	22.6	102.4	.	23.0	105.9	571.3	17.6	66.3	847.8
20-21.	22.3	107.9	642.5	21.5	100.1	555.5	19.9	114.9	985.5	18.9	106.0	.
22.	19.9	97.5	694.5	20.3	95.2	362.8	20.1	91.9	443.8	20.2	104.3	592.6
23.	19.0	99.4	743.9	19.7	101.2	576.3	18.6	97.1	439.0	17.9	90.9	2168.0
24.	20.9	105.7	1096.7	20.9	111.1	588.2	20.4	102.6	839.5	19.6	103.3	622.8
25.	18.1	90.9	429.6	18.8	91.9	428.6	17.7	85.8	433.8	18.7	94.7	346.7
26.	19.4	100.7	1031.7	21.6	102.4	551.9	19.4	101.4	915.8	18.4	116.2	925.5
27.	19.7	100.0	898.9	20.9	102.6	766.8	18.8	102.2	1047.7	19.3	79.0	612.0
28.	20.1	98.8	551.4	20.5	100.3	615.0	19.8	94.7	1034.3	19.3	88.0	408.5
29-30.	21.5	110.5	1417.6	21.0	102.8	1029.3	19.1	106.9	850.9	20.9	108.0	1785.7
31.	17.4	97.5	375.6	19.0	92.9	399.4	18.1	86.4	471.5	19.3	75.0	802.7
32.	18.0	97.7	320.5	19.4	95.3	2042.7	17.5	88.5	446.8	17.4	77.3	306.0
33.	17.1	95.3	1051.0	17.6	88.1	341.3	17.3	92.1	1381.3	17.9	81.3	376.8

Source: Data processing on ISTAT 2007 data

Table A1.2

σ by stratum (NUTS-1, firm size, economic sectors), Italy, 2007

NUTS1 and firm size / NACE Sectors	North-West			North-East			Centre			South		
	10-49	50-249	250+	10-49	50-249	250+	10-49	50-249	250+	10-49	50-249	250+
10-12.	9.3	49.2	951.4	9.6	52.1	1106.1	8.7	48.9	279.3	8.9	40.0	185.9
13.	10.6	43.6	228.3	9.3	44.7	372.8	8.6	37.2	130.9	10.3	46.0	1.5
14.	8.7	43.5	1087.3	7.9	41.6	273.8	8.9	40.5	200.1	9.0	37.6	397.2
15.	8.7	34.4	62.0	9.5	47.4	169.2	8.7	33.7	784.9	8.8	40.1	64.7
16-17.	9.3	45.0	918.2	9.1	44.9	159.3	8.8	40.4	122.0	8.1	60.2	294.9
18.	9.0	46.5	113.8	8.5	43.1	417.7	8.3	40.6	0.0	8.0	18.4	0.0
19.	11.8	58.5	369.0	9.8	41.0	.	12.1	60.5	352.3	7.9	23.9	429.8
20-21.	10.8	52.3	682.6	10.9	44.6	301.6	10.1	51.5	964.6	9.5	48.6	.
22.	10.1	48.0	922.4	10.0	44.9	145.6	10.0	43.0	74.7	10.3	49.5	393.5
23.	9.6	49.1	694.9	9.9	52.2	380.8	9.3	41.8	200.5	8.9	39.2	0.0
24.	10.4	50.1	2627.6	9.9	53.9	383.9	10.1	51.4	734.2	9.8	41.4	416.3
25.	9.0	43.0	183.5	9.3	44.7	165.3	8.8	38.7	322.2	9.4	43.9	104.4
26.	9.6	49.8	1816.3	10.8	52.7	270.9	9.5	49.5	887.2	9.6	52.2	690.7
27.	9.8	48.8	988.7	10.3	51.1	1318.6	9.3	53.7	1364.8	9.8	25.5	362.0
28.	10.1	45.7	458.0	10.1	47.9	539.5	10.1	46.0	1250.0	9.3	34.6	95.6
29-30.	10.7	52.0	2995.6	10.4	52.1	1552.9	9.5	52.5	1085.3	10.4	47.6	2210.7
31.	8.3	46.9	107.7	9.7	42.5	158.2	8.6	38.3	262.7	10.0	25.0	1035.9
32.	8.7	48.5	84.0	9.7	46.1	2441.8	8.7	39.1	179.9	8.2	19.0	0.0
33.	8.3	43.4	972.6	8.5	45.4	49.4	8.0	41.1	1249.6	8.6	35.6	139.4

Source: Data processing on ISTAT, 2007 data

Notes: '.' Empty stratum/no units allocated; $\bar{Y}$ = 35.0; 10-12. Manufacture of food, beverages and tobacco products; 13. Manufacture of textiles; 14. Manufacture of wearing apparel; 15. Manufacture of leather and related products; 16-17. Manufacture of wood and paper products; 18. Printing and reproduction of recorded media; 19. Manufacture of coke and refined petroleum products; 20-21. Manufacture of chemical and pharmaceutical products; 22. Manufacture of rubber and plastic products; 23. Manufacture of other non-metallic mineral products; 24. Manufacture of basic metals; 25. Manufacture of fabricated metal products, except machinery and equipment; 26. Manufacture of computer, electronic and optical products; 27. Manufacture of electrical equipment; 28. Manufacture of machinery and equipment n.e.c.; 29-30. Manufacture of transport vehicles; 31. Manufacture of furniture; 32. Other manufacturing; 33. Repair and installation of machinery and equipment.

A2. Appendix: Stratum Allocation: (Modified Proportional, Neyman, ROAUST(u=9), IPNLP)

Table A2.1
Proportional$_{(u=1)}$: Firms allocation by stratum (NUTS-1, firm size, economic sectors), Italy, 2007 (units)

NUTS1 and firm size / NACE Sectors	North-West			North-East			Centre			South			Total
	10-49	50-249	250+	10-49	50-249	250+	10-49	50-249	250+	10-49	50-249	250+	
10-12.	71	11	3	80	13	2	52	4	1	89	8	1	335
13.	68	15	3	25	4	1	44	4	1	13	2	1	181
14.	63	6	1	82	7	2	55	5	1	59	4	1	286
15.	14	3	1	40	6	1	94	7	1	29	3	1	200
16-17.	58	7	1	69	8	1	41	4	1	35	3	1	229
18.	43	4	1	34	3	1	23	2	1	14	1	1	128
19.	2	1	1	1	1	.	1	1	1	4	1	1	15
20-21.	30	14	4	16	6	1	11	3	2	11	2	.	100
22.	72	13	2	45	9	1	25	4	1	20	3	1	196
23.	42	6	1	54	11	3	40	5	1	56	5	1	225
24.	29	10	2	13	6	1	8	2	1	8	2	1	83
25.	288	29	2	215	20	2	93	8	1	94	10	1	763
26.	33	7	2	20	5	1	13	3	1	6	2	1	94
27.	48	10	2	38	8	2	16	3	1	9	2	1	140
28.	146	30	4	127	28	5	34	6	1	24	3	1	409
29-30.	26	9	4	17	5	2	15	4	1	11	4	1	99
31.	41	5	1	75	11	1	42	5	1	22	3	1	208
32.	27	4	1	31	5	1	22	2	1	9	1	1	105
33.	63	4	1	43	2	1	31	2	1	36	3	1	188
Total	1164	188	37	1025	158	29	660	74	20	549	62	18	3984

Source: Data processing on ISTAT, 2007 data

Table A2.2
Neyman: Firms allocation by stratum (NUTS-1, firm size, economic sectors), Italy, 2007 (units)

NUTS1 and firm size / NACE Sectors	North-West			North-East			Centre			South			Total
	10-49	50-249	250+	10-49	50-249	250+	10-49	50-249	250+	10-49	50-249	250+	
10-12.	36	28	54	42	36	39	25	10	10	43	17	7	347
13.	39	35	29	13	8	8	21	7	1	7	3	.	171
14.	30	14	22	35	14	17	26	10	4	29	8	6	215
15.	7	4	.	21	14	6	45	11	9	14	5	.	136
16-17.	29	16	20	34	18	5	19	8	3	16	7	3	178
18.	21	9	3	15	6	5	10	3	.	6	1	.	79
19.	1	1	6	.	1	.	1	1	3	1	.	4	19
20-21.	18	37	89	9	13	11	6	8	32	5	4	.	232
22.	40	34	41	25	21	6	14	9	1	11	7	7	216
23.	22	15	19	29	29	47	20	11	8	27	9	.	236
24.	16	25	42	7	15	15	4	4	8	4	3	5	148
25.	142	67	19	109	49	13	45	17	7	48	23	3	542
26.	17	18	29	11	13	10	6	7	13	3	3	4	134
27.	26	24	36	21	20	28	8	8	15	5	1	3	195
28.	81	75	88	70	72	105	19	14	8	12	5	1	550
29-30.	15	23	79	9	13	31	8	9	15	6	10	19	237
31.	18	10	1	40	24	7	20	9	5	12	3	6	155
32.	13	10	2	17	12	7	10	3	2	4	.	.	80
33.	29	9	5	20	5	.	13	3	4	17	6	2	113
Total	600	454	584	527	383	360	320	152	148	270	115	70	3983

Source: Data processing on ISTAT, 2007 data

Table A2.3
ROAUST$_{(u=9)}$: Firms allocation by stratum (NUTS-1, firm size, economic sectors), Italy, 2007 (units)

NUTS1 and firm size / NACE Sectors	North-West			North-East			Centre			South			Total
	10-49	50-249	250+	10-49	50-249	250+	10-49	50-249	250+	10-49	50-249	250+	
10-12.	27	23	54	30	27	39	21	14	14	31	18	12	310
13.	29	27	24	15	13	9	19	13	3	12	10	2	176
14.	24	16	22	27	16	18	22	14	8	24	13	6	210
15.	12	11	2	19	16	12	32	15	9	16	11	3	158
16-17.	24	17	20	26	18	12	19	13	10	17	13	4	193
18.	20	14	11	17	12	5	14	11	1	12	9	1	127
19.	9	10	7	9	5	.	9	7	4	10	6	4	80
20-21.	18	28	84	14	15	15	12	13	32	12	11	.	254
22.	29	26	41	21	19	12	16	14	4	15	13	7	217
23.	20	17	19	24	24	32	19	14	13	23	14	1	220
24.	17	21	42	12	17	16	11	11	8	11	11	5	182
25.	80	43	19	64	33	15	32	17	9	33	21	10	376
26.	18	18	29	15	16	14	12	12	13	10	10	4	171
27.	22	21	36	20	19	28	13	13	15	11	10	3	211
28.	49	46	59	44	45	79	18	16	8	15	11	4	394
29-30.	17	21	79	14	15	31	13	14	15	12	14	19	264
31.	18	14	5	29	21	13	19	14	8	15	11	6	173
32.	15	14	10	17	15	7	14	11	4	11	6	1	125
33.	23	14	5	19	11	3	16	10	4	17	12	6	140
Total	471	401	568	436	357	360	331	246	182	307	224	98	3981

Source: Data processing on ISTAT, 2007 data

Table A2.4
IPNLP: Firms allocation by stratum (NUTS-1, firm size, economic sectors), Italy, 2007 (units)

NUTS1 and firm size / NACE Sectors	North-West			North-East			Centre			South			Total
	10-49	50-249	250+	10-49	50-249	250+	10-49	50-249	250+	10-49	50-249	250+	
10-12.	26	20	54	31	26	39	18	15	8	32	13	7	289
13.	29	26	22	18	14	7	17	13	2	21	10	1	180
14.	22	14	22	26	14	13	20	14	5	21	13	5	189
15.	18	9	1	17	15	7	33	12	9	17	12	1	151
16-17.	22	15	20	25	14	5	17	13	3	16	19	3	172
18.	18	15	4	16	12	4	17	10	1	16	4	1	118
19.	13	6	5	8	4	.	10	5	4	12	4	3	74
20-21.	17	28	89	18	13	9	18	12	32	17	11	.	264
22.	30	25	41	18	15	7	18	14	1	18	13	6	206
23.	18	16	19	22	22	35	18	12	8	20	12	1	203
24.	18	18	42	16	15	11	16	12	8	17	9	4	186
25.	105	50	14	81	36	9	33	14	7	35	17	4	405
26.	18	16	29	18	17	8	17	13	13	17	9	4	179
27.	19	18	36	18	17	28	17	15	15	17	6	3	209
28.	60	55	73	52	53	103	19	15	8	17	10	2	467
29-30.	18	17	79	17	16	31	18	14	15	17	12	19	273
31.	17	14	3	29	18	7	17	13	6	19	7	6	156
32.	17	15	4	18	15	7	18	10	3	15	3	1	126
33.	21	13	5	17	14	1	16	10	4	17	12	4	134
Total	506	390	562	465	350	331	357	236	152	361	196	75	3981

Source: Data processing on ISTAT, 2007 data

Notes: '.' Empty stratum/no units allocated; 10-12. Manufacture of food, beverages and tobacco products; 13. Manufacture of textiles; 14. Manufacture of wearing apparel; 15. Manufacture of leather and related products; 16-17. Manufacture of wood and paper products; 18. Printing and reproduction of recorded media; 19. Manufacture of coke and refined petroleum products; 20-21. Manufacture of chemical and pharmaceutical products; 22. Manufacture of rubber and plastic products; 23. Manufacture of other non-metallic mineral products; 24. Manufacture of basic metals; 25. Manufacture of fabricated metal products, except machinery and equipment; 26. Manufacture of computer, electronic and optical products; 27. Manufacture of electrical equipment; 28. Manufacture of machinery and equipment n.e.c.; 29-30. Manufacture of transport vehicles; 31. Manufacture of furniture; 32. Other manufacturing; 33. Repair and installation of machinery and equipment.

A3. Appendix: Stratum RRMSE: (Modified Proportional, Neyman, ROAUST(u=9), IPNLP)

Table A3.1
Proportional$_{(u=1)}$: RRMSE by stratum (NUTS-1, firm size, economic sectors), Italy, 2007

NUTS1 and firm size / NACE Sectors	North-West			North-East			Centre			South		
	10-49	50-249	250+	10-49	50-249	250+	10-49	50-249	250+	10-49	50-249	250+
10-12.	0.0547	0.1456	0.7771	0.0546	0.1345	0.8967	0.0677	0.2443	0.5076	0.0521	0.1462	0.4137
13.	0.0617	0.1107	0.2707	0.0978	0.2261	0.6280	0.0703	0.2256	0.3184	0.1480	0.3296	0.0048
14.	0.0583	0.1906	1.3356	0.0491	0.1572	0.3873	0.0655	0.2085	0.4565	0.0640	0.2175	0.6296
15.	0.1296	0.2236	0.1877	0.0744	0.1957	0.4354	0.0491	0.1627	0.9954	0.0919	0.2770	0.1960
16-17.	0.0680	0.1721	1.2788	0.0558	0.1575	0.3221	0.0763	0.2177	0.2441	0.0761	0.3668	0.5369
18.	0.0746	0.2489	0.2670	0.0751	0.2546	0.7737	0.0992	0.3063	0.0000	0.1204	0.2726	0.0000
19.	0.3673	0.4366	0.5754	0.4490	0.3999	.	0.5404	0.5809	0.6110	0.2178	0.3636	0.5185
20-21.	0.0843	0.1311	0.5016	0.1254	0.1800	0.5369	0.1520	0.2566	0.6710	0.1461	0.3146	.
22.	0.0583	0.1374	0.9432	0.0732	0.1543	0.3985	0.0998	0.2295	0.1699	0.1103	0.2642	0.6637
23.	0.0774	0.1947	0.9165	0.0679	0.1526	0.3771	0.0733	0.1881	0.4582	0.0656	0.1890	0.0000
24.	0.0897	0.1438	1.6587	0.1289	0.1956	0.6539	0.1784	0.3533	0.8838	0.1759	0.2819	0.6706
25.	0.0287	0.0878	0.3061	0.0338	0.1057	0.2596	0.0485	0.1532	0.7406	0.0508	0.1470	0.3036
26.	0.0841	0.1842	1.2610	0.1101	0.2217	0.4860	0.1345	0.2923	0.9685	0.2134	0.3105	0.7488
27.	0.0703	0.1495	0.7727	0.0771	0.1717	1.1734	0.1178	0.2952	1.3250	0.1634	0.2244	0.6078
28.	0.0408	0.0816	0.3992	0.0426	0.0881	0.4000	0.0859	0.1967	1.1996	0.0978	0.2192	0.2377
29-30.	0.0967	0.1545	1.0155	0.1150	0.2268	1.0213	0.1325	0.2335	1.2659	0.1516	0.2122	1.2325
31.	0.0721	0.2155	0.2860	0.0559	0.1388	0.4063	0.0684	0.1924	0.5563	0.1040	0.1939	1.2881
32.	0.0919	0.2420	0.2589	0.0861	0.2169	1.2210	0.1051	0.3094	0.4024	0.1492	0.2456	0.0000
33.	0.0595	0.2204	0.9201	0.0694	0.3656	0.1456	0.0820	0.3020	0.8946	0.0782	0.2444	0.3726

Source: Data processing on ISTAT, 2007 data

Table A3.2
Neyman: RRMSE by stratum (NUTS-1, firm size, economic sectors), Italy, 2007

NUTS1 and firm size / NACE Sectors	North-West			North-East			Centre			South		
	10-49	50-249	250+	10-49	50-249	250+	10-49	50-249	250+	10-49	50-249	250+
10-12.	0.0781	0.0897	0.0000	0.0756	0.0767	0.0000	0.0987	0.1469	0.0927	0.0738	0.0997	0.1193
13.	0.0819	0.0689	0.0614	0.1378	0.1582	0.0770	0.1034	0.1643	0.3184	0.2103	0.2574	.
14.	0.0877	0.1188	0.0000	0.0735	0.1125	0.0801	0.0956	0.1422	0.1748	0.0925	0.1546	0.0000
15.	0.1844	0.1879	.	0.1077	0.1209	0.1435	0.0708	0.1245	0.0000	0.1274	0.2036	.
16-17.	0.0939	0.1125	0.0000	0.0838	0.1016	0.1154	0.1099	0.1508	0.1213	0.1128	0.2286	0.1786
18.	0.1110	0.1519	0.1386	0.1203	0.1743	0.0000	0.1532	0.2520	.	0.1891	0.2726	.
19.	0.5263	0.4366	0.0941	.	0.3999	.	0.5404	0.5809	0.1994	0.4394	.	0.0000
20-21.	0.1134	0.0758	0.0000	0.1638	0.1191	0.0856	0.2020	0.1495	0.0000	0.2235	0.2187	.
22.	0.0771	0.0816	0.0000	0.1000	0.0967	0.1323	0.1347	0.1501	0.1699	0.1471	0.1666	0.0000
23.	0.1082	0.1200	0.0000	0.0929	0.0931	0.0240	0.1084	0.1224	0.1179	0.0969	0.1365	.
24.	0.1214	0.0870	0.0000	0.1728	0.1202	0.0443	0.2490	0.2450	0.0000	0.2416	0.2280	0.0000
25.	0.0411	0.0546	0.0752	0.0470	0.0664	0.0809	0.0706	0.1037	0.1367	0.0737	0.0907	0.1519
26.	0.1163	0.1119	0.0000	0.1518	0.1335	0.0894	0.2006	0.1792	0.0000	0.3027	0.2426	0.0000
27.	0.1011	0.0954	0.0000	0.1043	0.1047	0.0000	0.1756	0.1731	0.0000	0.2223	0.3305	0.0000
28.	0.0564	0.0525	0.0000	0.0573	0.0525	0.0000	0.1160	0.1215	0.0000	0.1393	0.1674	0.2377
29-30.	0.1269	0.0904	0.0000	0.1610	0.1341	0.0000	0.1747	0.1514	0.0000	0.2068	0.1325	0.0000
31.	0.1117	0.1437	0.2860	0.0783	0.0891	0.1180	0.1052	0.1431	0.1611	0.1363	0.1939	0.0000
32.	0.1323	0.1476	0.1762	0.1194	0.1333	0.0000	0.1603	0.2549	0.2322	0.2356	.	.
33.	0.0858	0.1466	0.0000	0.1040	0.2241	.	0.1231	0.2505	0.0000	0.1162	0.1652	0.2318

Source: Data processing on ISTAT, 2007 data

Table A3.3
ROAUST$_{(u=9)}$: RRMSE by stratum (NUTS-1, firm size, economic sectors), Italy, 2007

NUTS1 and firm size / NACE Sectors	North-West			North-East			Centre			South		
	10-49	50-249	250+	10-49	50-249	250+	10-49	50-249	250+	10-49	50-249	250+
10-12.	0.0930	0.1014	0.0000	0.0920	0.0899	0.0000	0.1081	0.1159	0.0354	0.0897	0.0965	0.0576
13.	0.0933	0.0803	0.0733	0.1305	0.1177	0.0000	0.1073	0.1176	0.0000	0.1554	0.1187	0.0000
14.	0.0979	0.1092	0.0000	0.0842	0.1055	0.0747	0.1031	0.1166	0.0000	0.1016	0.1136	0.0000
15.	0.1414	0.1040	0.0000	0.1134	0.1121	0.0493	0.0844	0.1046	0.0000	0.1184	0.1259	0.0000
16-17.	0.1031	0.1082	0.0000	0.0973	0.1016	0.0373	0.1099	0.1171	0.0000	0.1097	0.1588	0.0000
18.	0.1131	0.1157	0.0000	0.1126	0.1180	0.0000	0.1302	0.1126	0.0000	0.1295	0.0683	0.0000
19.	0.1522	0.0000	0.0000	0.1012	0.0000	.	0.1379	0.0000	0.0000	0.1298	0.0000	0.0000
20-21.	0.1134	0.0903	0.0280	0.1329	0.1082	0.0000	0.1451	0.1114	0.0000	0.1423	0.1148	.
22.	0.0920	0.0949	0.0000	0.1080	0.1023	0.0607	0.1249	0.1171	0.0000	0.1269	0.1146	0.0000
23.	0.1135	0.1087	0.0000	0.1043	0.1033	0.0709	0.1112	0.1042	0.0597	0.1047	0.1049	0.0000
24.	0.1181	0.0973	0.0000	0.1330	0.1099	0.0000	0.1461	0.1214	0.0000	0.1472	0.0983	0.0000
25.	0.0550	0.0703	0.0752	0.0621	0.0827	0.0713	0.0855	0.1037	0.0000	0.0874	0.0953	0.0293
26.	0.1125	0.1119	0.0000	0.1296	0.1178	0.0353	0.1400	0.1280	0.0000	0.1583	0.1053	0.0000
27.	0.1102	0.1009	0.0000	0.1066	0.1080	0.0000	0.1314	0.1306	0.0000	0.1454	0.0784	0.0000
28.	0.0717	0.0669	0.0636	0.0741	0.0697	0.0503	0.1186	0.1112	0.0000	0.1227	0.1119	0.0000
29-30.	0.1180	0.0950	0.0000	0.1268	0.1226	0.0000	0.1412	0.1188	0.0000	0.1449	0.1077	0.0000
31.	0.1117	0.1171	0.0000	0.0921	0.0962	0.0622	0.1068	0.1120	0.0000	0.1229	0.0871	0.0000
32.	0.1253	0.1217	0.0365	0.1194	0.1216	0.0000	0.1306	0.1106	0.0000	0.1350	0.0000	0.0000
33.	0.0968	0.1139	0.0000	0.1068	0.1321	0.0000	0.1135	0.1086	0.0000	0.1162	0.1150	0.0000

Source: Data processing on ISTAT, 2007 data

Table A3.4
IPNLP: RRMSE by stratum (NUTS-1, firm size, economic sectors), Italy, 2007

NUTS1 and firm size / NACE Sectors	North-West			North-East			Centre			South		
	10-49	50-249	250+	10-49	50-249	250+	10-49	50-249	250+	10-49	50-249	250+
10-12.	0.0955	0.1078	0.0000	0.0904	0.0919	0.0000	0.1180	0.1122	0.1263	0.0876	0.1166	0.1193
13.	0.0933	0.0817	0.0793	0.1187	0.1138	0.1151	0.1130	0.1176	0.1615	0.1163	0.1187	0.0048
14.	0.1016	0.1188	0.0000	0.0851	0.1125	0.1095	0.1090	0.1166	0.1373	0.1087	0.1136	0.1257
15.	0.1143	0.1203	0.1877	0.1189	0.1155	0.1245	0.0840	0.1187	0.0000	0.1146	0.1176	0.1960
16-17.	0.1079	0.1155	0.0000	0.0990	0.1174	0.1154	0.1171	0.1171	0.1213	0.1128	0.1138	0.1786
18.	0.1219	0.1110	0.1088	0.1156	0.1180	0.1978	0.1170	0.1196	0.0000	0.1101	0.1257	0.0000
19.	0.1106	0.1178	0.1464	0.1117	0.1017	.	0.1233	0.1382	0.0000	0.1164	0.1137	0.1689
20-21.	0.1164	0.0903	0.0000	0.1173	0.1191	0.1150	0.1140	0.1169	0.0000	0.1194	0.1148	.
22.	0.0898	0.0966	0.0000	0.1170	0.1144	0.1164	0.1195	0.1171	0.1699	0.1167	0.1146	0.1112
23.	0.1196	0.1146	0.0000	0.1089	0.1090	0.0618	0.1164	0.1146	0.1179	0.1105	0.1143	0.0000
24.	0.1150	0.1063	0.0000	0.1155	0.1202	0.1138	0.1196	0.1147	0.0000	0.1153	0.1127	0.1613
25.	0.0476	0.0642	0.0931	0.0561	0.0782	0.1121	0.0837	0.1121	0.1367	0.0840	0.1099	0.1235
26.	0.1125	0.1203	0.0000	0.1174	0.1140	0.1206	0.1160	0.1185	0.0000	0.1162	0.1149	0.0000
27.	0.1180	0.1087	0.0000	0.1137	0.1152	0.0000	0.1142	0.1170	0.0000	0.1119	0.1131	0.0000
28.	0.0651	0.0615	0.0409	0.0671	0.0628	0.0122	0.1160	0.1171	0.0000	0.1136	0.1179	0.1346
29-30.	0.1157	0.1071	0.0000	0.1150	0.1185	0.0000	0.1207	0.1188	0.0000	0.1209	0.1179	0.0000
31.	0.1152	0.1171	0.1135	0.0921	0.1050	0.1180	0.1147	0.1175	0.1200	0.1136	0.1173	0.0000
32.	0.1186	0.1152	0.1098	0.1148	0.1216	0.0000	0.1175	0.1180	0.1266	0.1154	0.1098	0.0000
33.	0.1005	0.1205	0.0000	0.1143	0.1106	0.1456	0.1135	0.1086	0.0000	0.1162	0.1150	0.1164

Source: Data processing on ISTAT, 2007 data

Notes: '.' Empty stratum/no units allocated; Manufacture of food, beverages and tobacco products; 13. Manufacture of textiles; 14. Manufacture of wearing apparel; 15. Manufacture of leather and related products; 16-17. Manufacture of wood and paper products; 18. Printing and reproduction of recorded media; 19. Manufacture of coke and refined petroleum products; 20-21. Manufacture of chemical and pharmaceutical products; 22. Manufacture of rubber and plastic products; 23. Manufacture of other non-metallic mineral products; 24. Manufacture of basic metals; 25. Manufacture of fabricated metal products, except machinery and equipment; 26. Manufacture of computer, electronic and optical products; 27. Manufacture of electrical equipment; 28. Manufacture of machinery and equipment n.e.c.; 29-30. Manufacture of transport vehicles; 31. Manufacture of furniture; 32. Other manufacturing; 33. Repair and installation of machinery and equipment.