CONTRIBUTI ISTAT

n. 2/2007

Production of job vacancy statistics: coverage

C. Baldi, D. Bellisai, S. Fivizzani e M. Sorrentino

Istat

SISTEMA STATISTICO NAZIONALE
ISTITUTO NAZIONALE DI STATISTICA

Le collane esistenti presso l'ISTAT - *Rivista di Statistica Ufficiale*, *Contributi ISTAT* e *Documenti ISTAT* - costituiscono strumenti per promuovere e valorizzare l'attività di ricerca e per diffondere i risultati degli studi svolti, in materia di statistica ufficiale, all'interno dell' ISTAT, del SISTAN, o da studiosi esterni.

La *Rivista di Statistica Ufficiale* accoglie lavori che hanno come oggetto la misurazione dei fenomeni economici, sociali, demografici e ambientali, la costruzione di sistemi informativi e di indicatori, le questioni di natura metodologica, tecnologica o istituzionale connesse al funzionamento dei sistemi statistici e al perseguimento dei fini della statistica ufficiale.

I lavori pubblicati in *Contributi ISTAT* sono diffusi allo scopo di stimolare il dibattito intorno ai risultati preliminari di ricerca in corso.

I *Documenti ISTAT* forniscono indicazioni su linee, progressi e miglioramenti di prodotto e di processo che caratterizzano l'attività dell'Istituto.

Il Comitato di redazione esamina le proposte di lavori da pubblicare nelle tre collane sopra indicate. Quelli pubblicati nei Contributi ISTAT e nei Documenti ISTAT sono valutati preventivamente dai dirigenti dell'Istituto, mentre i lavori pubblicati nella Rivista di Statistica Ufficiale sono subordinati al giudizio di referee esterni.

ISTAT – Produzione libraria e centro stampa

CONTRIBUTI ISTAT

n. 2/2007

Production of job vacancy statistics: coverage

C. Baldi(*), D. Bellisai(*), S. Fivizzani(*) e M. Sorrentino(*)

(*) ISTAT - Servizio Statistiche congiunturali sull'occupazione e sui redditi

Contributi e Documenti Istat 2007

Istituto Nazionale di Statistica
Servizio Produzione Editoriale

Produzione libraria e centro stampa:
Carla Pecorario
Via Tuscolana, 1788 - 00173 Roma

AUTHORSHIP

This paper is the result of a joint work carried out by Ciro Baldi, Diego Bellisai, Stefania Fivizzani and Marina Sorrentino. However, section 2.1, 2.3 and 5 have been written by Stefania Fivizzani; sections 2.2, 6.2 and 6.3 have been written by Ciro Baldi; section 6.1 has been written by Diego Bellisai; section 3 has been written by Diego Bellisai and Marina Sorrentino; section 4 has been written by Diego Bellisai and Stefania Fivizzani; section 5.3 has been written by Ciro Baldi and Marina Sorrentino.

AKNOWLEDGMENTS

This paper is part of a project for the production of statistics on job vacancies, employment and hours worked. The authors express their gratitude to the other member of the project, Annalisa Lucarelli. Furthermore they wish to thank Leonello Tronti and Gianpaolo Oneto for their suggestions and feedbacks.

ABSTRACT

The upcoming EU regulation on Job Vacancies requires that the vacancy statistics cover the population of enterprises with at least one employee. At the moment the Italian quarterly survey on Job vacancies and Hours worked (VELA) covers enterprises with at least 10 employees. This report summarizes the results of a feasibility study, financed by Eurostat, aimed at extending the coverage below the 10 employee threshold. Since a direct method, involving the widening of the VELA sample, would imply a considerable increase of the costs of data collection, we investigate the possibility of using an indirect method involving auxiliary information to estimate the target variable for enterprises with less than 10 employees. This method uses the information on Job vacancies for enterprises with 10 to 20 employees and the information on employment on enterprises with less than 10 employees (from the OROS survey). In this report, given the availability of data from an experimental wave of the Excelsior Survey on hiring intentions (run by Unioncamere, the Union of Italian Chambers of Commerce, in collaboration with the Ministry of Labour and the European Social Fund) extended to include a section on job vacancies, we are able to fully test the abovementioned method. The results are encouraging, although more evidence is required to validate this kind of methods. Finally, an interesting by-product of this work is the attempt of estimating the number of job vacancies through a count data model.

CONTENTS

1. Introduction
2. The current information situation
    2.1 The Quarterly Job Vacancy Survey (VELA)
    2.2 The OROS survey
    2.3 The Excelsior Survey
3. An overview of possible methods to extend the coverage of the survey below 10 employees
4. Assessment of the error and descriptive statistics
5. Differences between the 1-9 and 10-19 size classes
    5.1 Inference
    5.2 Prediction
    5.3 Differences within enterprises with 1-9 employees
6. Modelling the job vacancies
    6.1 Choice of the response variable distribution
    6.2 Estimation and inference on a 1-9 size class effect
    6.3 Prediction of the JVR in enterprises 1-9 employees with a model estimated on those with 10-19 employees
7. Conclusions
References

LIST OF TABLES AND GRAPHS

Table 1 - Data sources: summary table

Table 2 - Job vacancy rate (JVR) - enterprises with at least 10 employees, differences with enterprises with at least one employee

Table 3 - Employees, job vacancies and job vacancy rate by size classes (level and percentage value)

Table 4 - Enterprises and employees by size class (level and percentage value)

Table 5 - Poisson model for job vacancies on enterprises with at least one employee

Table 6 - Poisson model for the job vacancy rate on enterprises with at least one employee

Table 7 - Job vacancy rate by economic activity and enterprise size

Table 8 - Frequency of enterprises with job vacancies by economic activity and enterprise size

Table 9 - Job vacancy rate on all enterprises with at least one employee - estimating 1-9 employees with 10-20 employees

Table 9a - Job vacancy rate by economic activity and enterprise size

Table 9b - Frequency of enterprises with job vacancies by economic activity and enterprise size

Table 10 - Poisson fitting of the distribution of job vacancies

Table 11 - Negative Binomial fitting of the distribution of job vacancies

Table 12 - Comparison of models: fit statistics

Table 13 - Negative binomial model for job vacancies on enterprises with 1-19 employees

Table 14 - Negative binomial model for the job vacancy rate on enterprises with 1-19 employees

Table 15 - Negative Binomial Regression model for job vacancy rate on enterprises with 10-20 employees

Table 16 - Job vacancy rate - enterprises with less than 10 employees: estimated vs. actual

Table 17 - Job vacancy rate - all enterprises with at least one employee: estimated vs. actual

Graph 1 - Poisson fitting of the distribution of job vacancies
Graph 2 - Negative Binomial fitting of the distribution of job vacancies

1. INTRODUCTION

Since the third quarter 2003, Istat has been running a quarterly business survey to produce estimates of job vacancies and occupied posts in order to comply with the gentlemen's agreement between Eurostat and the Member States.

At the moment, the target population of the survey includes all enterprises with at least 10 employees and with the activity classified in NACE Rev 1.1 sections C to K.

The aim of this report is to conduct a feasibility study to assess whether it is possible to extend the size coverage of the currently produced job vacancy statistics below the 10 employees threshold.

The extension of the sample below the threshold, with the aim of using a direct estimator to calculate job vacancy statistics, is considered unfeasible at least in the short and medium run given the number of enterprises that should be added and the cost that the operation would entail. However this strategy is not the only viable one. The availability of auxiliary variables from other sources opens the possibility of using indirect methods to estimate job vacancy statistics on the entire population of enterprises. Two data sources could be available to this aim. The first is the OROS survey database which provides quarterly information on employment on nearly the totality of enterprises with at least one employee. The second is the modified Excelsior survey on Hiring intentions run by Unioncamere (the Union of Italian Chambers of Commerce), in collaboration with the Ministry of Labour and the European Social Fund, which is a survey on enterprises with at least one employee. ISTAT is trying to set up a formal agreement with Unioncamere to add a small section to the questionnaire to collect data on job vacancies. In case of success we would have annual information on job vacancies on a (large) sample of enterprises including units currently excluded by the size threshold. In order to develop a pilot study, the additional section on job vacancies was added in the Excelsior 2005 wave (referring for job vacancies at the end of 2004)[2].

These are the data that are used in this report to perform the preliminary analysis aimed at the study of quality and reliability of indirect estimators. The possibility of adding questions on job vacancies to the Excelsior survey and the resulting availability of a very large sample of enterprises with less than 10 employees made unnecessary to launch a specific pilot survey to study this subpopulation. In other words, the job vacancy-enhanced Excelsior survey constitutes the source of information for the pilot study on small enterprises. It is worth pointing out that a separate entirely new survey run with the only purpose of the pilot study would have involved a much smaller number of respondents: in the project proposal, a sample of around 2,500 units was planned, to be compared with over 50,000 units in the Excelsior survey. On the other hand, the possibility of

[2] See ISTAT (2005b) and the interim technical report to the contract to which the present report refers to.

making use of a survey of an external institution, and hence the different strategy to perform this pilot study and, possibly, extend regularly the coverage of VELA, has required that the section on job vacancies, to be added to the basic Excelsior questionnaire, be as simple as possible and in line with the other sections. This requirement has barred the option of shaping the section on Job Vacancies differently for smaller and larger enterprises. In turn, these circumstances made useless a questionnaire design stage aimed at testing for the need for specific questionnaires, depending on the sample unit size, by studying hiring behaviours in smaller and larger enterprises However, informal discussions with persons responsible for the management of the Excelsior Survey have indicated that there seem to be no particular difficulties in answering the questionnaire depending on the size of the enterprise.

The report is structured as follows. Section 2 illustrates the available sources and their characteristics. Section 3 sketches the methods applied to extend the target coverage and describes the possible variants of indirect estimates. The following sections discuss the empirical analysis. In particular, in section 4 we give an assessment of the results obtained using the sample with the current threshold as an estimate for the entire population. In section 5 we present tests of the differences between the size class 1-9 and the neighbouring size class 10-19[3] and predictions of the job vacancy rate (in the following, JVR) worked out using the means of the neighbouring class as an estimate for the missing class. In Section 6 we extend the analysis of section 5 within a count data model framework. Some concluding remarks follow.

2. The current information situation

2.1 The Quarterly Job Vacancy Survey (VELA)

Istat has been running the quarterly business survey since the 3rd quarter 2003 in order to produce estimates of job vacancies and occupied posts for the population of all the enterprises with at least 10 employees and classified in section C to K of NACE Rev 1.1[4].

All aspects of the survey were designed so as to produce quarterly estimates of job vacancies and occupied posts only for this target population. The current sample size is about 8,000 units.

The reporting units receive a questionnaire each quarter asking the data needed to identify the enterprise, the number of employees, the number of employees who have started and stopped

[3] The size class is calculated on the OROS employment variable. This variable measures a quarterly average so that it can assume fractional values. When we refer to the size class 1-9, we are in reality talking about all enterprises that have strictly more than 0 employees and strictly less than 10 employees. Analogously the size class 10-19 refers to all enterprises with at least 10 employees and strictly less than 20 employees. In other terms, with the exception of the first class, the lower bound of the class is included and the upper bound represents the integer part of the maximum size in the class.

[4] For a detailed description of the survey, see ISTAT (2005a) and Chapter 1 (on timeliness) of the present final technical report.

working in the enterprise during the quarter, job vacancies at the end of the reference quarter, hours worked and hours paid but not worked.

Enterprises can transmit the data through the survey internet site, by e-mail or participating to the CATI.

2.2 The OROS survey

Since 2003 ISTAT has been releasing quarterly indicators on employment, wage and labour costs on the basis of the OROS survey[5]. This new survey is based on the use of the Social Security Institute (INPS) database of social contributions declarations, the DM10 forms, that have to be delivered monthly by enterprises with at least one employee in the private sector of the economy. This source is integrated, for enterprises with at least 500 employees, with the ISTAT monthly Survey on Large Enterprises which collects information on the same variables (and on the hours worked). The data contained in the DM10 forms refer to single months but they are aggregated by Istat and released with reference to quarters.

The survey at present covers the population of enterprises with at least one employee whose economic activity is in the sections C to K of the Nace Rev.1.1 classification.

Each quarter two sets of estimates are released: i) the "preliminary" estimate, based on a "non-random" sample of INPS data, with a delay of about 70 days from the end of the reference quarter; ii) a "final" estimate based on the "total population" of INPS data, with a delay of 15 months from the end of the reference quarter. The sample on which the preliminary estimate is based is the set of DM10 forms sent to INPS by electronic means. These forms are readily available since they are promptly uploaded from the local INPS offices to the central database. The need of a second estimate stemmed from the fact that the database of DM10 forms for a specific month took time to be fully populated as the local INPS offices had to manually enter the data delivered by enterprises on paper forms. However, a new national law (art. 44, Law n. 326/2003) requires that, starting from the reference month of January 2005, all DM10 forms have to be transmitted exclusively by electronic means with a maximum delay of one month. INPS anticipated the coming into force of this reform asking all enterprises to deliver the DM10 forms by electronic means from 2004. Moreover, from September 2004 the operators, namely banks and post offices, that had the task of receiving the DM10 forms were forbidden to accept paper forms from enterprises. Further details on the characteristics of this process can be found in ISTAT (2005c).

Since the end of 2004 the sample available for the preliminary estimates has grown steadily and now accounts for nearly 99% of the complete universe. The time necessary for the sample to coincide with the set of data on all units of the reference population remains still uncertain. To cope

[5] The employment estimates are still confidential and are not diffused at the national level.

with this near completeness of the set of electronic mode respondents a method to identify the late respondents and impute the data is currently being tested. Soon ISTAT could use data on all enterprises with at least one employee whose economic activity is in NACE Rev. 1.1 sections C to K to release current estimates on wages, labour costs and employment.

The availability of data on the entire population of enterprises at quarterly frequency has triggered a process of methodological re-thinking on the overall set of short term business statistics on labour market variables. The vacancy statistics will benefit from this extended data source in at least two ways: on the one hand the information on employment available at an enterprise level will be used to impute the employment variable for the VELA non respondents; on the other hand the estimate of total employment will be used to calibrate the VELA data so that the vacancy statistics will be referred to the current population of enterprises with at least 10 employees.

In the present context, the information on employment coming from OROS will be used to obtain auxiliary variables that can be used to predict the level of job vacancies for the subpopulation of enterprises with 1-9 employees.

The employment variable contained in the OROS data measures the quarterly average employment in each enterprise and, thus, it has a very strong correlation with the employment at the end of quarter measured by VELA.

2.3 The Excelsior Survey

Since 1997, Unioncamere (the Union of Italian Chambers of Commerce) has been carrying out a yearly business survey (named Excelsior), in order to measure hiring intentions in the following year by occupation, with the aim to supply a set of information to be used in choices on education and training by policymakers and by users at all levels.

In 2004 Istat and Unioncamere agreed to add to the questionnaire of the Excelsior survey for 2005 a set of additional questions aiming at measuring job vacancies by NACE Rev. 1.1 sections, NUTS 2 regions and ISCO 88 major occupational groups, as required by the gentlemen's agreement between Eurostat and the EU Member States. The aim of this experiment was to evaluate the effects of the addition on the survey and the quality of the data collected on job vacancies [6].

In particular, the purpose of this inclusion was twofold. On the one hand, as already mentioned, it aimed at assessing the feasibility of using this source to calculate annual job vacancy statistics by NACE Rev. 1.1 sections, NUTS 2 regions and ISCO 88 occupational major groups, as required by the gentlemen's agreement.

[6] For more details, see ISTAT(2005b).

On the other hand, it was meant to supply a database to be used in the assessment of whether the size coverage of the statistics on job vacancies currently produced by ISTAT could be extended below the 10 employees threshold, possibly via indirect estimators.

Hence, the modified version of the Excelsior survey including the job vacancy questions is intended in the present report as the pilot survey for the study on the size coverage extension.

The target population of the Excelsior survey includes all private enterprises with a least one employee present in the most recent release of the Chambers of Commerce business register (it excludes operational units of the public administration, public health sector enterprises, public schools and universities).

This report focuses on the survey on NACE sections C-K.

2.3.1 Sampling design and data collection issues

The survey sampled with probability strictly lower than one the population of enterprises with less than 50 employees, whilst it was exhaustive for larger enterprises.

The sampling design was based on the following stratification variables and classes:

- 43 economic activity sectors, based on aggregations of divisions and groups of the NACE classification[7];
- 5 size classes based on the number of employees: 1-9, 10-49, 50-99, 100-250, more than 250;
- 103 provinces (corresponding to the NUTS 3 geographical areas).

This stratification of the reference population identified 22,145 strata; in 8,858 of them the sampling was non exhaustive. However, many of them did not contain either any or enough population units. Hence, for each province Unioncamere chose to focus on the economic sectors in which the area is specialized. The strata identified by the other sectors were collapsed into larger ones.

The initial sample size was around 100,000 enterprises.

The allocation of the sample units in the strata had to satisfy a constraint on the maximum estimated standard error corresponding to a 95% significance level. In particular, it was imposed that this standard error could not be larger than 10% for two variables:

- a qualitative variable indicating whether an enterprise intended to hire or not in 2005;
- a quantitative variable measuring the hiring intentions.

The sampling design aimed at ensuring significant estimation results on 27 economic activity sectors.[8]

[7] Council Regulation (EEC) No 3037/90 of 9 October 1990.
[8] For a correspondence between this economic activity classification and the NACE, see Table 1in Istat (2005b).

The drawing of sample units and reserve pool was based on a further stratification of the above described strata. The sampling fraction was constant across sub-strata obtained from a given stratum and equal to that on the entire stratum (that is across sub-strata of a given stratum there was a proportional allocation of the sample units of the stratum).

The substitution of non responding units was based on minimum distance criteria, that weighted more heavily the proximity in terms of economic activity than in terms of geographical area.

The sample of respondents corresponded to 8.7% of the reference population, in terms of number of enterprises, but to a much greater share, 34.2%, in terms of employees.

The data collection was different on two subsets of the sample. The CATI technique was used for the 96,000 enterprises with up to 250 employees during the period December 2004-April 2005, while for the 3,200 largest enterprises face-to-face interviews were organized.

The survey collected data over a period of some months: this was a problem for the question about job vacancies. In fact, the nature of job vacancies characterized by volatility and the possibly informal way in which they leave track in the enterprise information system would suggest to prefer a data collection close to the reference date (in this case the last day of the year).

The Excelsior questionnaire is quite long and detailed. The additional questions on job vacancies were added in the section on the economic data concerning the occupations for which hiring intentions for 2005 were indicated. The aim of the additional questions is to measure job vacancies on 31 December 2004.

The job vacancy data have been collected only once within the Excelsior survey and at the moment their future collection has not yet been agreed upon.

An outline of the available data sources is presented in Table 1.

Table 1 - Data sources: summary table

Source of data	Periodicity	Coverage - NACE Sections	Coverage - Employer Size	Sample size	Main Variables	Reference time for Employment	Reference time for JV
Survey on Job vacancies and Hours Worked (VELA)	Quarter	C to K	at least 10 employees	8000	Employment, Job vacancies, Hours Worked	Beginning and End of quarter (point in time)	End of quarter
Survey on Employment, Wages and Labor Cost (OROS)	Quarter	C to K	at least 1 employees	1300000 (totality)	Employment, Wages, Labour Costs	Entire quarter (average)	-
Survey on Hiring Intentions (EXCELSIOR)	Year	A to O	at least 1 employees	100000	Employment, Hiring Intentions, Job vacancies	End of quarter (point in time)	End of quarter

The main source of information used in this report are the individual data on job vacancies and employment on 31 December 2004 from the Excelsior sample. These data were matched with the OROS data for the 4th quarter of 2004 in order to have auxiliary variables on each unit.

3. An overview of possible methods to extend the coverage of the survey below 10 employees

The information situation outlined in the previous sections has shown that two quarterly data sources are currently and regularly available: the Job Vacancy Survey on enterprises with at least 10 employees and the Survey on Employment, Wages and Labour Costs on enterprises with at least one employee. ISTAT is also trying to set up an agreement with Unioncamere to add a small section on Job vacancies to the Excelsior questionnaire and to get micro data on this and related sections of the survey. If this attempt were successful, the amount of information would become considerably larger.

In any case the availability of auxiliary information, that is information available from sources external to the main VELA survey, can be used to study an indirect method which extends the target population below 10 employees.

The problem we are facing can be seen as a small area estimation problem in the extreme case where no information is available on the variable of interest in the group of interest (the small area group). Hence, it can be useful to recall some terminology from the small area literature to place our problem in a more standard context.

Following Schaible (2000) a *direct estimator* can be defined as one which uses values of the variable of interest only from the reference period and only from units in the domain of interest, while an *indirect estimator* can be defined as one which uses values of the variable of interest from a domain and/or reference period other than the domain and time period of interest. A *time indirect estimator* is one which uses values of the variable of interest from another time period but the same domain of interest (i.e. it borrows strength from another time period). In contrast, a *domain indirect estimator* is one which uses values of the variable of interest from another domain but the same time period (i.e. it borrows strength from another domain).

In the present context, generally speaking three approaches to produce job vacancy data for enterprises with 1-9 employees can be envisaged:

1. a direct estimation method: the survey sample is extended to include this subpopulation of enterprises and the collected data are grossed up to a population including the very small enterprises;
2. an indirect estimation method without current information on the target variable: the survey sample is not extended and auxiliary information from external sources is used to produce estimates referring to all enterprises with at least one employee;
3. an indirect estimation method with current information on the target variable on a small set of units: a combination of an extension of the survey sample to enterprises smaller than the currently considered threshold and indirect estimation methods; the sample size below 10 employees should be smaller than in case 1.

The use of the direct estimation method is currently unfeasible. In fact, enterprises with 1-9 employees are around 86% of all enterprises with at least one employee, and employ around 27% of all employees (see Table 4). Furthermore, very small enterprises are characterised by high rates of birth, death and structural change. Hence, the number of respondents should be much larger than that of the currently run quarterly survey (around 8,000 units). Moreover, the number of enterprises to be contacted in order to achieve such a number of responses should be substantially larger. The costs to be faced for a similar survey in terms of human and financial resources cannot be met at the moment and neither in the medium term.

The indirect estimation method without extending the sample is the one considered in this report. In the framework of this method, three distinct analyses could be performed, where the job vacancy rate[9] for enterprises with 1-9 employees can be estimated using:

a. the rate for enterprises with 10-19 employees:
   - the basic assumption of this method is that, in appropriately chosen cells, the JVR of enterprises with less than 10 employees is not significantly different from the one for enterprises with 10-19 employees. This method could be applied quarterly to the data collected in the currently run survey. The estimate of the number of job vacancies could thus be obtained simply by multiplying the imputed JVR by an estimate of the end-of-quarter employment. The seasonal pattern of the JVR in the 1-9 employees size class would be similar to that of the 10-19 employees size class. The currently available information on the smaller enterprises does not allow to assess whether the two size classes actually show a similar seasonal pattern.
b. a model, linking the job vacancy rate (or job vacancies) to some measure of employment from an auxiliary source (OROS), estimated for enterprises with at least 10 employees and used to predict the dependent variable for those with 1-9 employees:
   - here the basic assumption is that, conditionally on the auxiliary variables, the JVR in enterprises with less than 10 employees is not significantly different from that in enterprises with 10-19 employees. This method too could be applied quarterly. It requires data on job vacancies collected in the survey currently run and data on employment from the OROS survey. It should produce a seasonal pattern in the 1-9 employees size class affected also by that of employment. Similarly to what indicated under method 1, the currently available information on the smaller enterprises does not allow us to assess whether the two variables display a similar seasonal pattern.

[9] Optimally the method should also estimate the employment at the end of the quarter that is the denominator of the Job Vacancy Rate. In what follows we make the assumption that the employment at the end of the quarter is perfectly estimated using the OROS data and focus on the estimation of the Job vacancy rate.

c. a model, linking the job vacancy rate (or job vacancies) to some measure of employment from an auxiliary source (OROS), estimated for enterprises with 1-9 employees once a year and used to predict the dependent variable for the same size class in all the subsequent quarters:
   - the basic assumption of this method is that the relation linking job vacancies and employment is stable within a year. This model can be estimated only for the 4th quarter of the year (at the moment, only for the 4th quarter of 2004), because it relies not only on data on employment from the OROS survey, but also on data on job vacancies in enterprises with 1-9 employees from the Excelsior survey. Furthermore, in this case, the seasonal characteristics of one quarter are imposed on all four.

In this report, we will try to assess the reliability of domain indirect estimators, that make use of the neighbouring group of enterprises with 10-19 employees, to estimate data on the 1-9 employees group not covered by the quarterly survey. We will talk interchangeably of estimation and imputation: this is because, given the presence of OROS information on employment on all enterprises, one can see the problem as one of imputation of JVRs and job vacancies on the 1-9 group (an item non response imputation problem).

The dataset we will use is defined by the Excelsior sample, with employees according to OROS. Thus we will be able to compare the predictions to the actual values. In other words, we will compare the actual values of the target variable in the 1-9 group to the estimates obtained by using the target variable in the neighbouring group and the auxiliary information in the 1-9 group.

We are not able to evaluate the performance of a time indirect estimator since we do not have actual job vacancy data in the 1-9 class for quarters other than the 4th one of 2004. However, one of the by-products of this work is to evaluate a model for job vacancies for the entire 1-19 size class, in which the sub-class 1-9 accounts for the vast majority of units. This is a pre-requisite for an estimation strategy based on a time indirect estimator.

4. Assessment of the error and descriptive statistics

The following tables present an analysis of the structure of job vacancies on the basis of the Excelsior sample[10]. Here and elsewhere, given the nature of pilot study of this work, the figures represent just parameters calculated on the sample data and no attempt is made to make a complete estimation with a grossing up to the universe. Job vacancy rates are calculated as percentage ratios

[10] For the sake of completeness, the data refer to the part of the Excelsior sample matched with the OROS universe in the 4th quarter of 2004, that is to 82,387 enterprises. Hence, this set covers the sectors from C to K (given the coverage of OROS) and does not include enterprises with employees according to Excelsior and with no employees or inactive according to OROS. However, since the OROS data refer to all enterprises in the reference population, the number of enterprises present in the Excelsior sample and lost in the matching is irrelevant.

between job vacancies and employees, both on 31 December 2004 and as measured by the Excelsior survey.

Table 2 - Job vacancy rate (JVR) - enterprises with at least 10 employees, differences with enterprises with at least one employee

NACE Section	JVR - enterprises with at least 10 employees	JVR - enterprises with at least one employee	JVR - difference between enterprises with one and with at least 10 employees
TOT	0.92	1.06	-0.14
C	0.77	1.27	-0.49
D	0.75	0.91	-0.16
E	0.39	0.43	-0.04
F	1.21	1.65	-0.44
G	1.38	1.48	-0.1
H	1.72	1.99	-0.27
I	0.66	0.72	-0.06
J	0.66	0.68	-0.02
K	1.32	1.41	-0.1

Source: Survey on Hiring Intentions (EXCELSIOR)

As a starting point, the error that would be committed if the estimates on the sample of enterprises with at least 10 employees were used to provide estimates for all enterprises with at least one employee is considered. To this end, Table 2 compares the JVR calculated only on the subpopulation of enterprises with at least 10 employees with that on the entire population of enterprises with at least one employee. The latter represents the target parameter, so that the difference between the first and the second can be considered as a measure of the estimation bias.

The error is systematically negative in all sections of economic activity, hinting at the fact that in smaller enterprises the job vacancy rate is higher than in larger ones. In the industry and services as a whole this error is equal to 0.14 percentage points but there are considerable differences among sectors. Values significantly higher than the average are observed in particular in C –Mining (-0.49), F –Constructions (-0.44), H -Hotels and Restaurants (-0.27).

Table 3 - Employees, job vacancies and job vacancy rate by size classes *(level and percentage value)*

Size class	N of employees - OROS	N of employees - EXCELSIOR	N of job vacancies	Job vacancy rate *(%)*
Total	3200325	3189714	33675	1.06
Up to 5	77849	93843	3420	3.64
5-10	105061	108910	2893	2.66
10-20	188336	190900	3244	1.7
20-50	257492	241686	3551	1.47
50-100	318896	294150	3349	1.14
100-500	881645	869972	7650	0.88
500-over	1371046	1390253	9568	0.69

Sources: Survey on Employment, Wage and Labor Cost (OROS); Survey on Hiring Intentions (EXCELSIOR)

Table 3 shows the JVRs by size classes. As implied by the previous results, rates higher than the average (1.06) are observed in smaller enterprises (3.64% in the enterprises with less than 5 employees and 2.66% in those with 5-9 employees), while in larger enterprises rates are smaller

than the average. Overall the JVR decreases with the size of enterprise. This is obviously due to the fact that the average number of job vacancies increases with the enterprise size but less than the number of employees does.

Table 4 - Enterprises and employees by size class (*level and percentage value)*

Size class	N of enterprises	% of enterprises	N of employees	% of employees
Total	1213136	100.0	10397149	100.0
Up to 5	862310	71.1	1602544	15.4
5-10	184266	15.2	1232040	11.8
10-20	99020	8.2	1326826	12.8
20-50	45018	3.7	1357587	13.1
50-100	12685	1	868907	8.4
100-500	8511	0.7	1640718	15.8
500-over	1326	0.1	2368527	22.8

Source: Survey on Employment, Wage and Labor Cost (OROS)

Although the analysis on job vacancies just reports sample averages, an illustration of the structure of the underlying population by enterprise size can be obtained from table 4. The figures, derived from the final data of the OROS survey, refer to the 4th quarter of 2004.

The share of the population of enterprises with 1-9 employees corresponds to 86% of the reference population in terms of number of enterprises and to 27% in terms of employees. On the other hand, enterprises with at least 100 persons employed represent a small share of all enterprises but almost 40% of total employees. These simple shares highlight once again the well known strong concentration of Italian enterprises in the segments of micro and small size.

The large share of employment in the 1-9 size class associated with a high (and above average) JVR leads us to imagine that the overall JVR would be considerably under-estimated if we used the sample of respondents with at least 10 employees to represent the whole population of enterprises with at least one employee.

The analysis of the JVR by size classes suggests that, to impute the values for the missing class, one should use data on the neighbouring class whose JVR is much closer in magnitude. However, since the difference in magnitude can depend from compositional effects rather than reflect the "intrinsic propensity" of enterprises of larger size to have lower job vacancy rates, the differences among size classes are also analyzed within a regression framework. To this end, Poisson regression models[11] are estimated on all enterprise sizes for the two possible dependent variables (job vacancies and the job vacancy rate on 31 December 2004), including as regressors both the explanatory variables based on the OROS employees data that a preliminary data analysis has identified as relevant (average number of employees in the 4th quarter 2004 and its squared

[11] Here we use a Poisson regression model since it is a baseline model for count data such as the number of Job vacancies. In section 6 a more complete analysis of the choice of the response variable distribution is performed.

value) and a large set of possible control variables. The considered control variables are: enterprise size (in 7 classes: 1-4, 5-9, 10-19, 20-49, 50-99, 100-499, 500+ employees); geographical area (in 4 classes: North-West, North-East, Centre, South including Sardinia and Sicily); economic activity (by NACE section for C and E-K, and by subsection for D).

Table 5 - Poisson model for job vacancies on enterprises with at least one employee

Parameter		Estimate	Standard Error	Chi-Square	Pr > ChiSq
Intercept		-2.9083	0.0961	915.08	<.0001
ditm404		0.0003	0.0000	2,642.39	<.0001
ditm404sq		0.0000	0.0000	780.39	<.0001
size	**2**	0.7426	0.0253	858.37	<.0001
size	**3**	0.9606	0.0246	1,519.72	<.0001
size	**4**	1.5838	0.0241	4,302.00	<.0001
size	**5**	2.1466	0.0247	7,561.93	<.0001
size	**6**	2.9647	0.0211	19,704.90	<.0001
size	**7**	4.2568	0.0240	31,435.50	<.0001
area	**2**	0.1378	0.0140	97.57	<.0001
area	**3**	0.1092	0.0158	47.64	<.0001
area	**4**	0.2019	0.0166	147.53	<.0001
nace1	**DA**	0.1998	0.0987	4.10	0.0429
nace1	**DB**	-0.2072	0.0994	4.34	0.0371
nace1	**DC**	-0.4622	0.1150	16.16	<.0001
nace1	**DD**	0.3949	0.1036	14.53	0.0001
nace1	**DE**	0.0596	0.1022	0.34	0.5596
nace1	**DF DG**	0.2182	0.0998	4.78	0.0288
nace1	**DH**	0.0559	0.1022	0.30	0.5841
nace1	**DI**	0.1639	0.1001	2.68	0.1016
nace1	**DJ**	0.2744	0.0965	8.08	0.0045
nace1	**DK**	0.2451	0.0969	6.39	0.0115
nace1	**DL**	0.1813	0.0984	3.40	0.0653
nace1	**DM**	0.1010	0.1014	0.99	0.3193
nace1	**DN**	0.1881	0.1003	3.52	0.0608
nace1	**E**	-0.2287	0.1159	3.89	0.0485
nace1	**F**	0.4268	0.0965	19.55	<.0001
nace1	**G**	0.6755	0.0951	50.40	<.0001
nace1	**H**	0.7937	0.0966	67.52	<.0001
nace1	**I**	0.5053	0.0963	27.56	<.0001
nace1	**J**	0.5196	0.0972	28.58	<.0001
nace1	**K**	0.5972	0.0953	39.27	<.0001
Scale		1	0		

Number of Observations Used: 82,387

Criteria For Assessing Goodness Of Fit			
Criterion	**DF**	**Value**	**Value/DF**
Deviance	8.20E+04	135180.3068	1.6414
Scaled Deviance	8.20E+04	135180.3068	1.6414
Pearson Chi-Square	8.20E+04	406054.9031	4.9305
Scaled Pearson X2	8.20E+04	406054.9031	4.9305
Log Likelihood		-24260.0194	

Table 6 - Poisson model for the job vacancy rate on enterprises with at least one employee

Parameter		Estimate	Standard Error	Chi-Square	Pr > ChiSq
Intercept		-3.8693	0.0961	1,620.67	<.0001
ditm404		-0.0001	0.0000	317.11	<.0001
ditm404sq		0.0000	0.0000	53.32	<.0001
size	**2**	-0.2496	0.0253	97.02	<.0001
size	**3**	-0.6860	0.0246	776.66	<.0001
size	**4**	-0.7780	0.0242	1,037.13	<.0001
size	**5**	-0.9735	0.0247	1,551.66	<.0001
size	**6**	-1.2362	0.0211	3,420.12	<.0001
size	**7**	-1.2140	0.0243	2,487.23	<.0001
area	**2**	0.2000	0.0140	204.80	<.0001
area	**3**	0.0271	0.0160	2.85	0.0913
area	**4**	0.3054	0.0166	338.61	<.0001
nace1	**DA**	0.2161	0.0987	4.79	0.0286
nace1	**DB**	-0.2521	0.0994	6.43	0.0112
nace1	**DC**	-0.5109	0.1150	19.74	<.0001
nace1	**DD**	0.4054	0.1036	15.31	<.0001
nace1	**DE**	-0.0274	0.1022	0.07	0.7884
nace1	**DF DG**	0.1863	0.0998	3.49	0.0619
nace1	**DH**	0.0141	0.1022	0.02	0.8905
nace1	**DI**	0.0991	0.1001	0.98	0.3222
nace1	**DJ**	0.2261	0.0965	5.48	0.0192
nace1	**DK**	0.1337	0.0969	1.90	0.1677
nace1	**DL**	0.1275	0.0984	1.68	0.1952
nace1	**DM**	0.0528	0.1014	0.27	0.6025
nace1	**DN**	0.1567	0.1004	2.44	0.1183
nace1	**E**	-0.3335	0.1159	8.28	0.004
nace1	**F**	0.4217	0.0965	19.08	<.0001
nace1	**G**	0.6371	0.0951	44.83	<.0001
nace1	**H**	0.8020	0.0966	68.95	<.0001
nace1	**I**	0.1812	0.0965	3.53	0.0604
nace1	**J**	0.4054	0.0973	17.35	<.0001
nace1	**K**	0.6241	0.0953	42.88	<.0001
Scale		1	0		

Number of Observations Used: 80,395

Criteria For Assessing Goodness Of Fit			
Criterion	**DF**	**Value**	**Value/DF**
Deviance	8.00E+04	136750.4595	1.7017
Scaled Deviance	8.00E+04	136750.4595	1.7017
Pearson Chi-Square	8.00E+04	415192.7376	5.1665
Scaled Pearson X2	8.00E+04	415192.7376	5.1665
Log Likelihood		-25045.0957	

Note: The difference between the number of observations used in table 5 and table 6 is due to the cases where the number of employees in an enterprise which responded to Excelsior was equal to zero on 31 December 2004, because this variable is the denominator of the JVR.

It can be observed (Table 5 for the model for job vacancies, and Table 6 for the model for the job vacancy rate) that the number of employees in the quarter preceding the job vacancy reference date and its squared value are both significant regressors. Furthermore, in these models, in general, both economic activity and geographical area significantly affect job vacancies and the job vacancy rate. The result to be emphasized here is that, even controlling for the number of employees, area and economic activity, an inverse relationship remains between the job vacancy rate and the size of the enterprise (a direct relationship when the dependent variable is the number of job vacancies).

Since the larger the enterprise the more its behaviour is different from that of enterprises with 1-9 employees, these results confirm that it is advisable to focus only on units with 10-19 employees to estimate job vacancies on those with 1-9 employees.

5. Differences between the 1-9 and 10-19 size classes

In the remainder of the paper we will try to assess whether information derived from the 10-19 size class can be used to impute job vacancies on the 1-9 size class. This section, in particular, is devoted to evaluate the differences on job vacancy statistics between the two size classes. We will perform two kinds of analysis: an inferential test to study differences between the means and a prediction analysis to measure the error we would incur in if we used the means on job vacancies on the larger size class to impute values on the smaller one.

5.1 Inference

The following tables compare the means of the job vacancy rate and the share of enterprises with at least one job vacancy between two groups: enterprises with 1-9 employees and enterprises with 10-19 employees.

Table 7 - Job vacancy rate by economic activity and enterprise size

NACE Section	Firm Size		F test on equality of variances		T test on difference between firm size classes Variances			
					Equal		Unequal	
	Up to 10	10-20						
	%	%	F value	Pr > |F|	t value	Pr > |t|	t value	Pr > |t|
C	3.23	1.97	1.5	0	1.26	0.21	1.39	0.17
D	3.97	1.85	4.1	0	11.39	0	15.77	0
E	3.52	1.28	8.39	0	1.01	0.32	1.6	0.11
F	4.78	2.64	2.46	0	4.75	0	5.9	0
G	3.45	2.04	3.51	0	4.37	0	6.02	0
H	4.23	3.6	1.52	0	0.94	0.35	1.09	0.28
I	4.05	3.88	1.07	0.27	0.24	0.81	0.25	0.81
J	3.33	5.28	1.66	0	-1.15	0.25	-0.93	0.36
K	3.59	4.18	1.03	0.51	-1.16	0.24	-1.15	0.25

Source: Survey on Hiring Intentions (EXCELSIOR)

Table 7 confirms, as already noted above, that job vacancy rates are higher in smaller enterprises. Two notable exceptions are the service sectors J and K. The last part of the table reports on the t-test on the differences. One should use this test under the hypothesis of equal or different

variances according to the test of equality of variances (shown in the fourth and fifth column). As long as the two variances are statistically different, the difference of means test to be used is the one reported in the last two columns of the table. We can conclude that the difference in means is not statistically significant in almost all the sections of economic activity.

Table 8 - Frequency of enterprises with job vacancies by economic activity and enterprise size

NACE Section	Firm Size		F test on equality of variances		T test on difference between firm size classes Variances			
					Equal		Unequal	
	Up to 10	10-20						
	%	%	F value	Pr > |F|	t value	Pr > |t|	t value	Pr > |t|
C	11.48	14.09	1.2	0.17	-0.84	0.4	-0.8	0.42
D	9.9	13.59	1.32	0	-8.73	0	-8.14	0
E	11.86	8.82	1.27	0.43	0.49	0.62	0.53	0.6
F	11.53	13.49	1.14	0	-2.09	0.04	-2.02	0.04
G	8.25	14.86	1.67	0	-9.27	0	-8.06	0
H	8.26	14.49	1.64	0	-5.25	0	-4.41	0
I	9.38	15.8	1.57	0	-4.93	0	-4.36	0
J	6.03	14.17	2.16	0	-3.39	0	-2.49	0.01
K	7.33	15.27	1.91	0	-9.19	0	-7.3	0

Source: Survey on Hiring Intentions

Results change if we look at the frequency of enterprises with job vacancies (table 8): in this case the probability for enterprises with 10-19 employees to have job vacancies is much higher than for smaller units. Inference supports the hypothesis that in general these differences are statistically relevant.

5.2. Prediction

The inferential analysis suggests that the JVR of enterprises with 10-19 employees can be attributed to those with less than 10 employees in the imputation phase. The next step aims precisely at measuring the error implied by the estimate of indicators by section based on this method.

Table 9 - Job vacancy rate on all enteprises with at least one employee - estimating 1-9 employees with 10-20 employees

NACE Section	Estimated	Actual	Difference
TOT	1.02	1.06	-0.04
C	1.05	1.27	-0.22
D	0.83	0.91	-0.08
E	0.40	0.43	-0.03
F	1.42	1.65	-0.23
G	1.43	1.48	-0.05
H	1.97	1.99	-0.02
I	0.75	0.72	0.03
J	0.72	0.68	0.03
K	1.51	1.41	0.10

Table 9 shows that the overall remaining error is markedly smaller than the one in table 2 and very small in absolute value. Moreover, setting aside C (whose small size could justify a larger error to be acceptable) and F, the error in each section is below 0.1 percentage points. Evidently,

caution must be used when interpreting this result for at least two reasons. First, this evaluation refers to only one point in time. Unfortunately there is no information available to replicate the analysis on other quarters or other years. Thus, in order to generalize this result one should assume that the implicit hypothesis that the JVR for enterprises with 10-19 employees is an unbiased (or, fairly unbiased) estimator of the JVR for enterprises with 1-9 employees is not affected by time factors (seasonal, cyclical, trend). Second, had this estimation strategy been carried out every quarter, the quality of the estimates could be quite different because the sample size of the regular ISTAT job vacancy survey is far smaller than the Excelsior one. Therefore, the mean squared error of this kind of estimate would be higher. Third, we are not considering relevant differences in the share of enterprises with job vacancies.

5.3 Differences within enterprises with 1-9 employees

In the following paragraphs, a set of experiments aimed at improving the prediction errors using a regression imputation strategy, instead of the mean imputation just described, are presented and discussed. Before turning to it, however, it is worth focusing on the analysis of the differences within the population with less than 10 employees. This is useful since the population of enterprises with 5-9 employees is far smaller than the entire population with 1-9 employees. Hence, one could consider the possibility of sampling from this subpopulation if the information that is drawn from it could also be used to better estimate the target variable on the smaller enterprises for which sampling would be too large and costly. Evidently, the efficiency gain of this approach, as an alternative to the method reported in paragraph 5.2 where in principle no new data have to be collected, rapidly vanishes if considerable differences on the two subpopulations are observed. With regard to tables 3 and 6, it was already remarked that the job vacancy rate is decreasing when the size increases and that significant differences between the two smallest subpopulations emerged.

Table 9a - Job vacancy rate by economic activity and enterprise size

NACE Section	Firm Size		F test on equality of variances		T test on difference between firm size classes Variances			
					Equal		Unequal	
	1-4	5-9						
	%	%	F value	Pr > |F|	t value	Pr > |t|	t value	Pr > |t|
C	3.50	2.95	1.50	0	0.51	0.61	0.51	0.61
D	4.52	2.96	2.86	0	7.40	0	8.53	0
E	4.09	2.88	5.54	0	0.51	0.61	0.53	0.6
F	5.31	3.76	2.77	0	3.06	0	3.55	0
G	3.66	2.81	3.03	0	2.24	0.03	2.89	0
H	4.45	3.56	1.95	0	1.47	0.14	1.73	0.08
I	4.54	3.04	2.20	0	2.24	0.03	2.54	0.01
J	3.30	3.50	1.60	0	-0.15	0.88	-0.18	0.86
K	3.60	3.55	1.49	0	0.12	0.91	0.13	0.89

Source: Survey on Hiring Intentions (EXCELSIOR)

Here the type of analysis used in paragraph 5.1 is applied in order to provide a breakdown by economic activity. Table 9a confirms that in most sections the JVR is higher for smaller enterprises and that the differences are highly significant from an inferential point of view.

Table 9b - Frequency of enterprises with job vacancies by economic activity and enterprise size

NACE Section	Firm Size		F test on equality of variances		T test on difference between firm size classes Variances			
					Equal		Unequal	
	1-4	5-9						
	%	%	F value	Pr > |F|	t value	Pr > |t|	t value	Pr > |t|
C	9.39	13.66	1.39	0.02	-1.37	0.17	-1.36	0.17
D	8.43	12.58	1.42	0	-9.95	0	-9.45	0
E	7.69	16.98	1.99	0.01	-1.55	0.12	-1.5	0.14
F	9.43	15.54	1.54	0	-6.32	0	-5.91	0
G	6.63	13.02	1.83	0	-8.89	0	-7.68	0
H	7.41	10.93	1.42	0	-3.54	0	-3.24	0
I	8.51	11.10	1.27	0	-2.11	0.03	-2.03	0.04
J	5.01	11.73	2.19	0	-3.49	0	-2.68	0.01
K	6.10	12.28	1.88	0	-8.16	0	-6.77	0

Source: Survey on Hiring Intentions (EXCELSIOR)

Furthermore, the frequencies of enterprises with vacancies, reported in table 9b, are significantly higher in the 5-9 employees size class in all sections but C and E.

All in all, the analysis summarised in this paragraph suggests that it is not worth expanding the sample to cover the population with 5-9 employees.

6. Modelling the job vacancies

As a further attempt to reduce the imputation error we try to use estimated parameters on a model for job vacancies estimated on enterprises with 10-19 employees to predict the values of the interest variable on enterprises with less than 10 employees. The effort of modelling job vacancies is also useful in order to explore the relationship with employment and other variables with the aim of using a model estimated once a year to predict the interest variable in other quarters.

The analysis is divided into the following steps:

1. choice of the response variable distribution;
2. estimation and inference of the effect of the size class 1-9 in a count data model estimated on all enterprises with 1-19 employees;
3. prediction of JVRs for enterprises with 1-9 employees with a model estimated on those with 10-19 employees.

6.1. Choice of the response variable distribution

The first step in our modelling analysis consists in choosing a distribution for the response variable. We have tried to fit some selected distributions, in particular the Poisson and the negative binomial ones to the actual job vacancy frequency distribution.

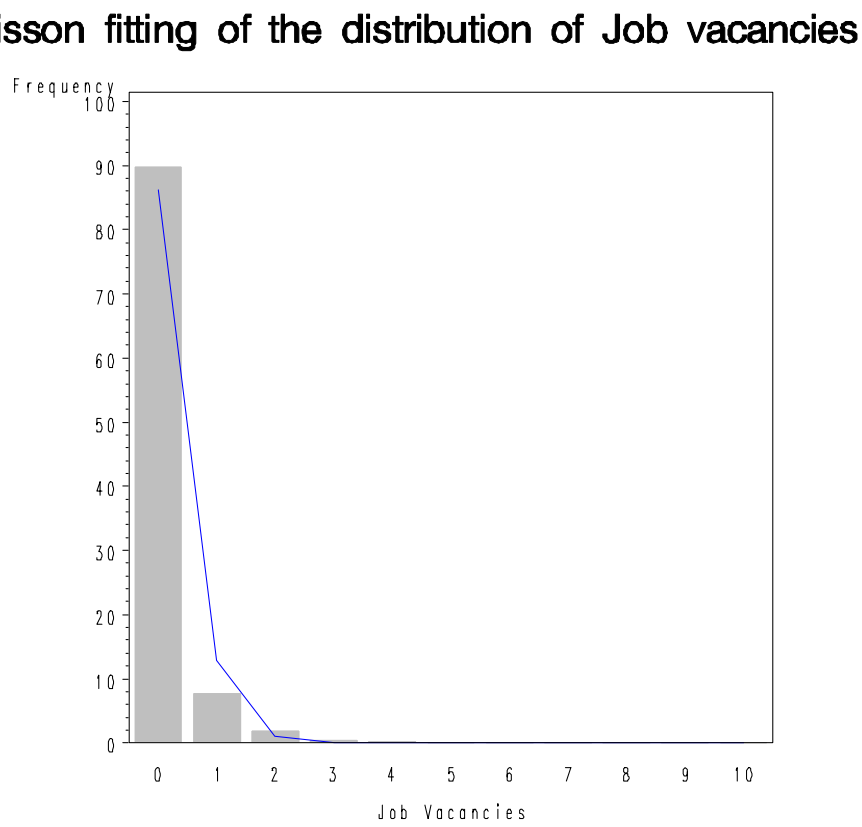
We have chosen the two abovementioned theoretical distributions since job vacancies can be seen as an event count variable, i.e. the realization of a non-negative integer-valued random variable.

The Poisson probability distribution, whose functional form is

$$f(y;\mu) = \frac{\mu^y e^{-\mu}}{y!}$$

has mean and variance equal to μ.

Graph 1. – Poisson Fitting of the Distribution of Job Vacancies



In graph 1, along with the actual frequency of enterprises with the specified number of vacancies, the probability to have that value of vacancies as generated by a Poisson distribution with mean equal to the sample mean and the Poisson-predicted number of enterprises for each observed value of job vacancies are shown. As one can immediately see, the Poisson distribution markedly underestimates all the count data, apart from the case of enterprises with one vacancy where it quite strongly overestimates the frequency of enterprises. The fit is evaluated by the Pearson's χ^2 and the likelihood ratio G-test reported in table 10. Both tests reject the hypothesis that a Poisson distribution adequately describes the data.

Table 10 - Poisson fitting of the distribution of job vacancies

Goodness-of-fit test

Analysis variable: Job Vacancy
Distribution: POISSON
Estimated Parameters: lambda = 0.1485

Pearson chi-square = 4.4373669E66
Prob > chi-square = 0

Likelihood ratio G2 = 8552.0153123
Prob > chi-square = 0

Degrees of freedom = 20

The second theoretical probability distribution considered is the negative binomial one, which is characterized by two parameters and is most commonly parameterised as

$$g(y;\mu,\nu)=\frac{\Gamma(\nu+y)}{y!\Gamma(\nu)}\left(\frac{\nu}{\nu+\mu}\right)^{\nu}\left(\frac{\mu}{\nu+\mu}\right)^{y}$$

where $\mu,\nu>0$, $y\in\{0,1,2,...\}$, $\alpha = 1/\nu$ determines the degree of dispersion and Γ is Euler's gamma function, defined as

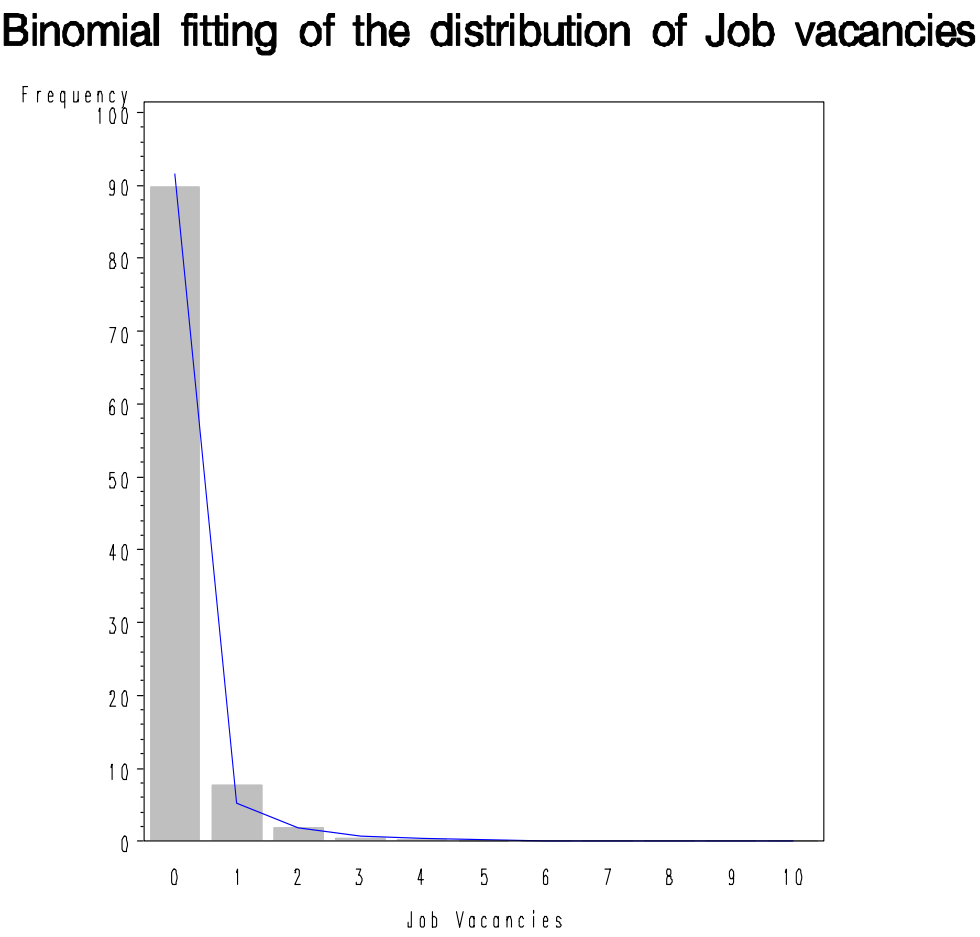
$$\Gamma(z)=\int_0^{\infty}t^{z-1}e^{-t}\mathrm{d}t\,.$$

The distribution mean and variance are given by

$$\mathrm{E}(y)=\mu\,,\qquad Var=\mu\left(1+\frac{\mu}{\nu}\right).$$

Hence, the variance of the negative binomial distribution increases with the dispersion parameter α.

Graph 2. – Negative Binomial Fitting of the Distribution of Job Vacancies



2.

The fitting is shown in graph 2 and table 11. It can be observed that the negative binomial probability distribution slightly overestimates the frequency of enterprises with zero and more than two vacancies, while it underestimates that of enterprises with one or two vacancies.

In this case too, the fit statistics do not support the hypothesis that the actual job vacancy distribution belongs to the considered family of theoretical ones. However, comparing the values of the Pearson's χ^2 and the likelihood ratio G-test for both fitted probability distributions, the negative binomial distribution seems slightly more suited to describe the data.

The quality of the fit can be improved if the mean is allowed to depend on some auxiliary variables. This is the reason why we have been led to investigate possible regression models based on these distributions.

For example, a Poisson regression model attributes to a response variable Y a Poisson distribution whose expected value depends on N regressors x_i in the following way:

$$\log(\mu) = b_0 + \sum_{i=1}^{N} b_i x_i$$

where $\{b_0, ..., b_N\}$ are coefficients to be estimated by a method such as the maximum likelihood.

We have thus explored four possible regression models, the first two related to the Poisson and the negative binomial distributions, and the other two related to the zero-inflated version of the two abovementioned models.

We chose to study also the zero-inflated Poisson (ZIP) and zero-inflated negative binomial (ZINB) regression models since they handle the phenomenon of over-dispersion[12] by changing the mean structure to explicitly model the presence of zero counts. The presence of large zero counts is precisely one of the main features of the job vacancy variable.

In order to achieve a zero-inflated model one chooses a model for the probability of having a specified count, e.g. a Poisson model, and "inflates" the zero count prediction by introducing an additional a parameter p_0 to rise the probability of zero count given by the Poisson model.

In this way the ZIP probability distribution can be written as

$$f_{ZIP}(y;\mu,p_0)=\begin{cases}\dfrac{(1-p_0)\mu^y e^{-\mu}}{y!}, y>0\\ p_0+(1-p_0)e^{-\mu}, y=0\end{cases},$$

where p_0 can be seen as a function of some auxiliary variables and is usually modelled with a logit or a probit regression.

Table 11 - Negative Binomial fitting of the distribution of job vacancies

Goodness-of-fit test

Analysis variable: Job Vacancy
Distribution: NEGBIN
Estimated Parameters: mu=0.1484; nu=0.0918

Pearson chi-square = 167151.29235
Prob > chi-square = 0

Likelihood ratio G2 = 1183.7406206
Prob > chi-square = 0

Degrees of freedom = 19

Analogous features characterizes the ZINB probability distribution.

To investigate possible regression models describing the relationship between job vacancies and employees in enterprises with less than 20 employees, we have considered, as regressors for the

[12] Overdispersion occurs when the observed variance is higher than the variance of a theoretical model. This circumstance is often encountered when fitting very simple parametric models, such as those based on the Poisson distribution. In fact, the Poisson distribution has one free parameter and does not allow for the variance to be adjusted independently from the mean. If there is overdispersion of the actual data with respect to the theoretical distribution, an alternative model with additional free parameters may provide a better fit.

mean: the number of employees on the 4th quarter 2004, its squared value and a set of dummy variables related to the enterprise geographical area and its economic activity.

As far as the ZIP and the ZINB regression models are concerned, p_0 is modelled and estimated with a logit regression using as regressors the number of employees on the 4th quarter 2004, its squared value and dummy variables representing the enterprise geographical area.

To determine which of the four abovementioned models better fits the job vacancy data for enterprises with less than 20 employees, some fit statistics have been calculated: the AIC, AICC and BIC indicators, as well as -2 times the log-likelihood.

The AIC (Akaike information criterion) indicator is a statistical model fit measure defined by the following formula

$AIC = 2k - 2\ln(L)$

where L is the likelihood function and k the number of regressors.

The AICC indicator is a variation of the AIC one to cope with small size samples and its formula is

$AICc = -2\ln(L) + 2kn/(n - k - 1)$

where n is the number of observations. The very large value of n in the present exercise implies that the sample values of the AIC and AICC indicators almost coincide.

Finally, the BIC (Bayesian information criterion) indicator is defined as

$BIC = -2\ln(L) + k\ln(n)$.

Hence all these fitting statistics impose a penalty for including too many regressors in the model.

Due to the functional forms of the criteria, a better fit of the data by the regression model is signalled by a lower value of the indicator.

Table 12 - Comparison of models: fit statistics

Model	Statistics			
	-2 Log Likelihood	AIC (smaller is better)	AICC (smaller is better)	BIC (smaller is better)
NegBin	47467.35	47521.35	47521.38	47764.63
Poisson	51571.39	51623.39	51623.42	51857.66
ZINB	49376.97	49442.97	49443.00	49740.30
ZIP	47942.85	48006.85	48006.88	48295.17

As can be noted from table 12, all considered criteria indicate that the best fit is obtained by the negative binomial regression model and the ZIP regression model, with a slight preference for the first one.

6.2. Estimation and inference on a 1-9 size class effect

The next step has been to assess the suitability of the negative binomial model (selected on the basis of the just described evidence) to predict job vacancies or the job vacancy rate in enterprises with 1-9 employees on the basis of its estimates on enterprises with 10-19 employees.

To this end, a model including a dummy variable for enterprises with 1-9 employees and its interactions with the number of employees in the 4th quarter 2004 and its squared value is estimated, both for job vacancies (table 13) and the job vacancy rate (table 14). Furthermore, geographical area and economic activity (with the same classes used for the Poisson models on all enterprise sizes for job vacancies and the job vacancy rate) are used as control variables.

Table 13 - Negative binomial model for job vacancies on enterprises with 1-19 employees

Parameter		Estimate	Standard Error	Chi-Square	Pr > ChiSq
Intercept		-1.2976	0.7208	3.24	0.0718
ditm404		-0.1123	0.1008	1.24	0.2651
ditm404sq		0.0054	0.0035	2.39	0.1218
d 1size1		-2.3463	0.7082	10.97	0.0009
ditm404size1		0.5356	0.1047	26.17	<.0001
ditm404sqsize1		-0.0313	0.0045	48.76	<.0001
area	**2**	0.1280	0.0397	10.40	0.0013
area	**3**	0.1742	0.0426	16.69	<.0001
area	**4**	0.4236	0.0388	118.94	<.0001
nace1	**DA**	0.0465	0.1584	0.09	0.7692
nace1	**DB**	-0.1585	0.1593	0.99	0.3197
nace1	**DC**	-0.5453	0.1846	8.73	0.0031
nace1	**DD**	0.2989	0.1623	3.39	0.0656
nace1	**DE**	0.0630	0.1647	0.15	0.7019
nace1	**DF DG**	0.0812	0.1849	0.19	0.6607
nace1	**DH**	0.1813	0.1690	1.15	0.2835
nace1	**DI**	0.0308	0.1619	0.04	0.8494
nace1	**DJ**	0.2457	0.1507	2.66	0.1030
nace1	**DK**	0.3044	0.1572	3.75	0.0528
nace1	**DL**	0.1105	0.1592	0.48	0.4875
nace1	**DM**	0.2804	0.2132	1.73	0.1885
nace1	**DN**	0.0711	0.1602	0.20	0.6572
nace1	**E**	0.2698	0.3172	0.72	0.3949
nace1	**F**	0.2915	0.1496	3.80	0.0514
nace1	**G**	0.0155	0.1486	0.01	0.9169
nace1	**H**	0.2096	0.1533	1.87	0.1714
nace1	**I**	0.1046	0.1571	0.44	0.5055
nace1	**J**	-0.2414	0.1908	1.60	0.2060
nace1	**K**	0.0652	0.1500	0.19	0.6641
Dispersion		3.7316	0.1183		

Number of Observations Used: 60,483

Criteria For Assessing Goodness Of Fit			
Criterion	**DF**	**Value**	**Value/DF**
Deviance	6.00E+04	21213.3505	0.3509
Scaled Deviance	6.00E+04	21213.3505	0.3509
Pearson Chi-	6.00E+04	60558.3754	1.0017
Scaled Pearson	6.00E+04	60558.3754	1.0017
Log Likelihood		-21545.889	

Table 14 - Negative binomial model for the job vacancy rate on enterprises with 1-19 employees

Parameter		Estimate	Standard Error	Chi-Square	Pr > ChiSq
Intercept		-2.4216	0.7211	11.28	0.0008
ditm404		-0.2582	0.1008	6.56	0.0104
ditm404sq		0.0081	0.0035	5.44	0.0197
d 1size1		-0.8856	0.7087	1.56	0.2114
ditm404size1		0.1340	0.1049	1.63	0.2012
ditm404sqsize1		-0.0038	0.0045	0.71	0.4000
area	**2**	0.1399	0.0398	12.38	0.0004
area	**3**	0.1874	0.0427	19.23	<.0001
area	**4**	0.4225	0.0390	117.56	<.0001
nace1	**DA**	0.0572	0.1586	0.13	0.7186
nace1	**DB**	-0.1555	0.1594	0.95	0.3294
nace1	**DC**	-0.5471	0.1848	8.76	0.0031
nace1	**DD**	0.3082	0.1625	3.60	0.0578
nace1	**DE**	0.0536	0.1648	0.11	0.7451
nace1	**DF DG**	0.0866	0.1850	0.22	0.6398
nace1	**DH**	0.1751	0.1691	1.07	0.3006
nace1	**DI**	0.0377	0.1620	0.05	0.8162
nace1	**DJ**	0.2403	0.1508	2.54	0.1110
nace1	**DK**	0.2853	0.1573	3.29	0.0697
nace1	**DL**	0.1079	0.1593	0.46	0.4983
nace1	**DM**	0.2454	0.2132	1.32	0.2498
nace1	**DN**	0.0669	0.1604	0.17	0.6767
nace1	**E**	0.2237	0.3186	0.49	0.4826
nace1	**F**	0.2992	0.1497	3.99	0.0457
nace1	**G**	0.0228	0.1487	0.02	0.8779
nace1	**H**	0.2490	0.1534	2.64	0.1045
nace1	**I**	0.1075	0.1572	0.47	0.4940
nace1	**J**	-0.2414	0.1913	1.59	0.2070
nace1	**K**	0.0588	0.1502	0.15	0.6954
Dispersion		3.7013	0.1179		

Number of Observations Used: 60,483

Criteria For Assessing Goodness Of Fit			
Criterion	**DF**	**Value**	**Value/DF**
Deviance	6.00E+04	21258.3764	0.3516
Scaled Deviance	6.00E+04	21258.3764	0.3516
Pearson Chi-Square	6.00E+04	60421.6271	0.9995
Scaled Pearson X2	6.00E+04	60421.6271	0.9995
Log Likelihood		-21538.851	

It can be observed that none of the three variables calculated using the dummy representing the 1-9 size class are significant regressors in the model for the job vacancy rate, while all of them appear as significant in the model for job vacancies. Hence, there is evidence that a common model for the rate can be estimated for enterprises with 1-9 and 10-19 employees. In other words, the

inferential results suggests that the relationship between the JVR and the employment variables does not depend on the size class.

The estimates of the negative binomial model for the job vacancy rate also indicate that both the average number of employees in the 4th quarter 2004 according to the OROS survey and its squared value significantly contribute to predict the dependent variable, together with the geographical area and economic activity (although for just a sub-set of considered classes). The evidence on the relevance of the control variables is similar in the estimates of the model for job vacancies, while in this model the two employment variables do not appear as significant for the 10-19 employees enterprises.

6.3 Prediction of the JVR in enterprises 1-9 employees with a model estimated on those with 10-19 employees

A second, more straightforward, way to evaluate the use of the specific model is to compare the parameter estimated on the basis of the values of the endogenous variable predicted by the model to the (actual) value of the parameter.

To this aim, first, a regression model for the job vacancy rate is estimated on the population of enterprises with 10-19 employees. Second, the value of the estimated coefficient is retained and used to predict the job vacancy rate on enterprises with 1-9 employees. Third, the average job vacancy rate calculated on the predicted values is compared to the (true) rate estimated on the actual values.

Table 15 - Negative Binomial Regression model for job vacancy rate on enterprises with 10-20 employees

Parameter		Estimate	Standard Error	Chi-Square	Pr > ChiSq
Intercept		-2.5712	0.8082	10.12	0.0015
ditm404		-0.2580	0.1082	5.69	0.0171
ditm404sq		0.0081	0.0038	4.68	0.0305
area	**2**	0.1038	0.0738	1.98	0.1599
area	**3**	0.0676	0.0811	0.70	0.4043
area	**4**	0.4011	0.0764	27.57	<.0001
nace1	**DA**	0.0899	0.3156	0.08	0.7758
nace1	**DB**	-0.0312	0.3009	0.01	0.9173
nace1	**DC**	-0.5307	0.3442	2.38	0.1232
nace1	**DD**	0.1515	0.3247	0.22	0.6408
nace1	**DE**	0.0952	0.3188	0.09	0.7653
nace1	**DF DG**	0.3739	0.3426	1.19	0.2752
nace1	**DH**	0.3401	0.3177	1.15	0.2844
nace1	**DI**	0.0953	0.3134	0.09	0.761
nace1	**DJ**	0.3473	0.2887	1.45	0.229
nace1	**DK**	0.4116	0.2989	1.90	0.1685
nace1	**DL**	0.3812	0.3049	1.56	0.2112
nace1	**DM**	0.4462	0.3791	1.39	0.2392
nace1	**DN**	0.1188	0.3074	0.15	0.6991
nace1	**E**	0.1858	0.6681	0.08	0.7809
nace1	**F**	0.4114	0.2873	2.05	0.1521
nace1	**G**	0.3016	0.2846	1.12	0.2893
nace1	**H**	0.6906	0.3016	5.24	0.0221
nace1	**I**	0.3960	0.3016	1.72	0.1893
nace1	**J**	0.0451	0.4426	0.01	0.9188
nace1	**K**	0.5243	0.2913	3.24	0.0719
Dispersion		4.8251	0.2246		

Number of Observations Used: 12,953

Criteria For Assessing Goodness Of Fit			
Criterion	**DF**	**Value**	**Value/DF**
Deviance	1.30E+04	5446.7003	0.4213
Scaled Deviance	1.30E+04	5446.7003	0.4213
Pearson Chi-Square	1.30E+04	14269.7072	1.1039
Scaled Pearson X2	1.30E+04	14269.7072	1.1039
Log Likelihood		-5910.3427	

Both the distribution of the dependent variable and the specification of the model for the mean follow the previous analysis. The estimate of the negative binomial regression model on the sample of enterprises with 10-19 employees is reported in table 15.

Table 16 - Job vacancy rate - enterprises with less than 10 employees: estimated vs. actual

NACE Section	Estimated	Actual	Difference
TOT	4.73	3.48	1.24
C	2.84	3.12	-0.27
D	3.86	3.41	0.45
E	3.51	3.15	0.36
F	4.90	4.16	0.74
G	4.91	3.11	1.79
H	7.23	3.74	3.50
I	5.00	3.52	1.48
J	4.25	3.13	1.12
K	6.44	3.53	2.90

The values of the coefficient estimates are used to predict those of the job vacancy rate on the sample of enterprises with less than 10 employees. Table 16 compares the results of the average value of the actual job vacancy rates to the average value of the predicted job vacancy rates for the target subsample. The difference between the estimated and the calculated JVR, shown in the fourth column, is a measure of the prediction error associated with the model. The error is systematically positive with higher values in the services sectors. We can compare these results with those in table 7 where the difference between the JVRs for enterprises with less than 10 employees and those with 10-19 employees is implicitly an estimate of the error that would be incurred in if the JVRs were estimated on the sample of larger size enterprises. As a whole, the model reduces the absolute value of the error in sectors C to F and rises it notably in the services sectors.

Table 17 - Job vacancy rate - all enterprises with at least one employee: estimated vs. actual

NACE Section	Estimated	Actual	Difference
TOT	1.17	1.06	0.12
C	1.28	1.27	0.01
D	0.97	0.91	0.06
E	0.43	0.43	0.00
F	1.81	1.65	0.15
G	1.66	1.48	0.18
H	2.48	1.99	0.49
I	0.80	0.72	0.08
J	0.71	0.68	0.02
K	1.67	1.41	0.26

In table 17 the overall bias on the entire coverage is reassessed by estimating the JVR of the enterprises with less than 10 employees with the values predicted by the model. With the remarkable exceptions of H and K, the size of the errors is very small. For H, in particular, one can conclude that the job vacancy behaviour of very small and larger enterprises is significantly different. However the overall error is 0.12 percentage points, thus somewhat smaller, in absolute

value, than the error in table 2, but markedly larger than that in table 9. Therefore, using a regression technique to model the job vacancy rate and to impute it on enterprises with less than 10 employees seems not to improve the results that can be obtained by a simple imputation of the mean from the neighbouring subgroup. Specification errors, but possibly also the weak correlation between job vacancies and employment, can be at the origin of this result.

7. Conclusions

In this report we have described the results of a preliminary analysis on the extension of the coverage of the current quarterly survey on job vacancies (VELA) to the enterprises with 1-9 employees in NACE rev. 1.1 sections C to K. Hence, these results have to be considered as a starting point for further analyses.

The main data source used in this analysis is the section on job vacancies of the Unioncamere Excelsior survey 2005 wave. These data were matched with information on employment coming from the OROS survey. This sample represents all the enterprises with at least one employee. This circumstance has allowed us to compare predicted and actual data.

The analysis has shown that the use of a sample of enterprises with at least 10 employees to estimate the JVR of the entire population generates a non-negligible bias. We have thus shown the improvements that can be obtained via indirect estimators that use the information of the subgroup with 10-19 employees to estimate/impute the JVR for the 1-9 employee subgroup.

In particular, an imputation-of-the-mean-like method produces a fairly good result even if the imputation cell is quite large (section level).

In contrast, the use of a regression imputation-like method does not reduce much the overall error and actually worsens the bias in some sections.

Furthermore, the significant differences pointed out by the relevant indicators between the 1-4 and 5-9 employees size classes support the choice not to extend the sampling to the 5-9 size class to indirectly estimate the data for the 1-4 class.

All the results must be considered with caution for two reasons. First, they are based on an only one point in time analysis. Second, the prediction error if VELA instead of Excelsior had been used could possibly be larger, since the VELA sample size is far smaller.

An alternative strategy has also been outlined in the report: if an annual data source on job vacancies on enterprises with less than 10 employees were available, it would be possible to estimate a relationship between job vacancies and auxiliary employment variables once a year and predict each quarter the target variable using the current information on the auxiliary employment variables from OROS. Obviously, this strategy would rely on the assumption that the modelled relationship is time invariant.

Further analyses are thus needed to judge the quality of an indirect method to extend the coverage of the vacancies survey. An important contribution in terms of empirical evidences would be provided by the replication of the analysis described in this report when new waves of the vacancy-enhanced Excelsior survey will be available. In this situation, the stability and the size of the prediction error of the indirect method could be studied, in the case both of mean imputation and regression imputation.

The availability of information on job vacancies for the subpopulation of enterprises with less than 10 employees repeated for two or more survey waves would also allow to assess the above mentioned alternative strategy. In fact, the parameters of a regression model estimated on data collected on the population of small enterprises in wave t could be used to predict job vacancies on this subpopulation for the period $t+s$. The prediction error could then be measured comparing the indirectly obtained estimates with those that can be worked out from the data actually collected on the subpopulation at time $t+s$.

Both approaches, being applied to data referred only to the last quarter of the year, would fail to supply evidence on the applicability of the methods to the other quarters. The absence of information on small enterprises for the first, second, and third quarter makes *de facto* impossible to fully test the methodologies. However, one piece of indirect supporting evidence would be provided by the study of the stability across quarters of the parameters of the regression model on the population of 10-19 employees estimated on the data collected in the current VELA survey.

Finally, a study should investigate the difference in the capability of the model to predict the Excelsior data if it was estimated on VELA data, that is on a far smaller sample, instead of on Excelsior ones.

This type of study, however, should also include an analysis of the difference between the vacancy variables collected by the two surveys. In fact, in the presence of systematic errors (due, for example, to slight differences in the questionnaire, or in the training of the CATI interviewers of the two surveys), the prediction error would be spurious being a mix of the actual prediction error due to model mis-specification and the error due to different measured variables. In case of systematic differences, the indirect methods based on Excelsior data should be modified to compensate for them. This, in turn, would involve a further stage of modelling and, hence, of analysis of suitable models.

In much of what has been just outlined it is clear the importance of the availability of an (at least) annual source of information on vacancies for small enterprises, like the vacancy-enhanced Excelsior survey. A final advantage would be represented by the possibility of using this source to

benchmark once a year the level of job vacancies estimated in each quarter via a domain indirect estimator similar to one of those used in sections 5 and 6.

The analyses performed in the project have provided for the first time in Italy empirical evidence concerning the measurement of the job vacancy variable for very small enterprises. However the results are still preliminary and must be carefully assessed, along the lines briefly sketched above, before taking practical decisions on whether and how data covering also enterprises with 1-9 employees can be regularly produced.

References

ISTAT (2005a), "Contract n. 2003 32100 022 - Launching and implementing the job vacancy statistics – Final Technical Report", May 2005.

ISTAT (2005b), "Contract n. 2003 32100 023 - Projects with regional dimension on methodology and data collection in the domain of job vacancies and the labour force survey – Final technical report", July 2005.

ISTAT (2005c) Grant agreement between the European Community and Istituto Nazionale di Statistica Title: "Hours worked, Wages & Salaries for Annex C, D" Interim Report 2005, 14 July

Schaible, Wes (2000), Methods for producing world and regional estimates for selected key indicators of the labour market. *ILO Employment paper n. 2000/6*, available at http://www.ilo.org/public/english/employment/strat/publ/ep00-6.htm#3

Contributi ISTAT(*)

1/2002 - Francesca Biancani, Andrea Carone, Rita Pistacchio e Giuseppina Ruocco - *Analisi delle imprese individuali*

2/2002 - Massimiliano Borgese .-.*Proposte metodologiche per un progetto d'indagine sul trasporto aereo alla luce della recente normativa comunitaria sul settore*

3/2002 - Nadia Di Veroli e Roberta Rizzi - *Proposta di classificazione dei rapporti di lavoro subordinato e delle attività di lavoro autonomo: analisi del quadro normativo*

4/2002 – Roberto Gismondi – *Uno stimatore ottimale in presenza di non risposte*

5/2002 – Maria Anna Pennucci – *Le strategie europee per l'occupazione dal Libro bianco di Delors al Consiglio Europeo di Cardiff*

1/2003 – Giovanni Maria Merola – *Safety Rules in Statistical Disclosure Control for Tabular Data*

2/2003 – Fabio Bacchini, Pietro Gennari e Roberto Iannaccone – *A new index of production for the construction sector based on input data*

3/2003 – Fulvia Cerroni e Enrica Morganti – *La metodologia e il potenziale informativo dell'archivio sui gruppi di impresa: primi risultati*

4/2003 – Sara Mastrovita e Isabella Siciliani – *Effetti dei trasferimenti sociali sulla distribuzione del reddito nei Paesi dell'Unione europea: un'analisi dal Panel europeo sule famiglie*

5/2003 – Patrizia Cella, Giuseppe Garofalo, Adriano Paggiaro, Nicola Torelli e Caterina Viviano – *Demografia d'impresa: l'utilizzo di tecniche di abbinamento per l'analisi della continuità*

6/2003 – Enrico Grande e Orietta Luzi – *Metodologie per l'imputazione delle mancate risposte parziali: analisi critica e soluzioni disponibili in Istat*

7/2003 – Stefania Fivizzani, Annalisa Lucarelli e Marina Sorrentino – *Indagine sperimentale sui posti di lavoro vacanti*

8/2003 – Mario Adua – *L'agricoltura di montagna: le aziende delle donne, caratteristiche agricole e socio-rurali*

9/2003 – Franco Mostacci e Roberto Sabbatini – *L'euro ha creato inflazione? Changeover e arrotondamenti dei prezzi al consumo in Italia nel 2002*

10/2003 – Leonello Tronti – *Problemi e prospettive di riforma del sistema pensionistico*

11/2003 – Roberto Gismondi – *Tecniche di stima e condizioni di coerenza per indagini infraannuali ripetute nel tempo*

12/2003 – Antonio Frenda – *Analisi delle legislazioni e delle prassi contabili relative ai gruppi di imprese nei paesi dell'Unione Europea*

1/2004 – Marcello D'Orazio, Marco Di Zio e Mauro Scanu – *Statistical Matching and the Likelihood Principle: Uncertainty and Logical Constraints*

2/2004 – Giovanna Brancato – *Metodologie e stime dell'errore di risposta. Una sperimentazione di reintervista telefonica*

3/2004 – Franco Mostacci, Giuseppina Natale e Elisabetta Pugliese – *Gli indici dei prezzi al consumo per sub popolazioni*

4/2004 – Leonello Tronti – *Una proposta di metodo: osservazioni e raccomandazioni sulla definizione e la classificazione di alcune variabili attinenti al mercato del lavoro*

5/2004 – Ugo Guarnera – *Alcuni metodi di imputazione delle mancate risposte parziali per dati quantitativi: il softaware* Quis

6/2004 – Patrizia Giaquinto, Marco Landriscina e Daniela Pagliuca – *La nuova funzione di analisi dei modelli implementata in Genesees v. 3.0*

7/2004 – Roberto Di Giuseppe, Patrizia Giaquinto e Daniela Pagliuca – *MAUSS (Multivariate Allocation of Units in Sampling Surveys):*

un software generalizzato per risolvere il problema dell' allocazione campionaria nelle indagini Istat

8/2004 – Ennio Fortunato e Liana Verzicco – *Problemi di rilevazione e integrazione della condizione professionale nelle indagini sociali dell'Istat*

9/2004 – Claudio Pauselli e Claudia Rinaldelli – *La valutazione dell'errore di campionamento delle stime di povertà relativa secondo la tecnica Replicazioni Bilanciate Ripetute*

10/2004 – Eugenio Arcidiacono, Marina Briolini, Paolo Giuberti, Marco Ricci, Giovanni Sacchini e Giorgia Telloli – *Procedimenti giudiziari, reati, indagati e vittime in Emilia-Romagna nel 2002: un'analisi territoriale sulla base dei procedimenti iscritti nel sistema informativo Re.Ge.*

11/2004 – Enrico Grande e Orietta Luzi **–** *Regression trees in the context of imputation of item non-response: an experimental application on business data*

12/2004 – Luisa Frova e Marilena Pappagallo **–** *Procedura di now-cast dei dati di mortalità per causa*

13/2004 – Giorgio DellaRocca, Marco Di Zio, Orietta Luzi, Emanuela Scavalli e Giorgia Simeoni **–** *IDEA (Indices for Data Editing Assessment): sistema per la valutazione degli effetti di procedure di controllo e correzione dei dati e per il calcolo degli indicatori SIDI*

14/2004 **–** Monica Pace, Silvia Bruzzone, Luisa Frova e Marilena Pappagallo **–** *Review of the existing information about death certification practices, certificate structures and training tools for certification of causes of death in Europe*

15/2004 – Elisa Berntsen – *Modello Unico di Dichiarazione ambientale: una fonte amministrativa per l'Archivio delle Unità Locali di Asia*

16/2004 – Salvatore F. Allegra e Alessandro La Rocca – *Sintetizzare misure elementari: una sperimentazione di alcuni criteri per la definizione di un indice composto*

17/2004 – Francesca R. Pogelli – *Un'applicazione del modello "Country Product Dummy" per un'analisi territoriale dei prezzi*

18/2004 – Antonia Manzari – *Valutazione comparativa di alcuni metodi di imputazione singola delle mancate risposte parziali per dati quantitativi*

19/2004 – Claudio Pauselli – *Intensità di povertà relativa: stima dell'errore di campionamento e sua valutazione temporale*

20/2004 – Maria Dimitri, Ersilia Di Pietro, Alessandra Nuccitelli e Evelina Paluzzi – *Sperimentazione di una metodologia per il controllo della qualità di dati anagrafici*

21/2004 – Tiziana Pichiorri, Anna M. Sgamba e Valerio Papale – *Un modello di ottimizzazione per l'imputazione delle mancate risposte statistiche nell'indagine sui trasporti marittimi dell'Istat*

(*) ultimi cinque anni

22/2004 – Diego Bellisai, Piero D. Falorsi, Annalisa Lucarelli, Maria A. Pennucci e Leonello G. Tronti – *Indagine pilota sulle retribuzioni di fatto nel pubblico impiego*

23/2004 – Lidia Brondi – *La riorganizzazione del sistema idrico: quadro normativo, delimitazione degli ambiti territoriali ottimali e analisi statistica delle loro caratteristiche strutturali*

24/2004 – Roberto Gismondi e Laura De Sandro – *Provisional Estimation of the Italian Monthly Retail Trade Index*

25/2004 – Annamaria Urbano, Claudia Brunini e Alessandra Chessa – *I minori in stato di abbandono: analisi del fenomeno e studio di una nuova prospettiva d'indagine*

26/2004 – Paola Anzini e Anna Ciammola – *La destagionalizzazione degli indici della produzione industriale: un confronto tra approccio diretto e indiretto*

27/2004 – Alessandro La Rocca – *Analisi della struttura settoriale dell'occupazione regionale: 8° Censimento dell'industria e dei servizi 2001 7° Censimento dell'industria e dei servizi 1991*

28/2004 – Vincenzo Spinelli e Massimiliano Tancioni – *I Trattamenti Monetari non Pensionistici: approccio computazionale e risultati della sperimentazione sugli archivi INPS-DM10*

29/2004 – Paolo Consolini – *L'indagine sperimentale sull'archivio fiscale modd.770 anno 1999: analisi della qualità del dato e stime campionarie*

1/2005 – Fabrizio M. Arosio – *La stampa periodica e l'informazione on-line: risultati dell'indagine pilota sui quotidiani on-line*

2/2005 – Marco Di Zio, Ugo Guarnera e Orietta Luzi – *Improving the effectiveness of a probabilistic editing strategy for business data*

3/2005 – Diego Moretti e Claudia Rinaldelli – *EU-SILC complex indicators: the implementation of variance estimation*

4/2005 – Fabio Bacchini, Roberto Iannaccone e Edoardo Otranto – *L'imputazione delle mancate risposte in presenza di dati longitudinali: un'applicazione ai permessi di costruzione*

5/2005 – Marco Broccoli – *Analisi della criminalità a livello comunale: metodologie innovative*

6/2005 – Claudia De Vitiis, Loredana Di Consiglio e Stefano Falorsi – *Studio del disegno campionario per la nuova rilevazione continua sulle Forze di Lavoro*

7/2005 – Edoardo Otranto e Roberto Iannaccone – *Continuous Time Models to Extract a Signal in Presence of Irregular Surveys*

8/2005 – Cosima Mero e Adriano Pareto – *Analisi e sintesi degli indicatori di qualità dell'attività di rilevazione nelle indagini campionarie sulle famiglie*

9/2005 – Filippo Oropallo – *Enterprise microsimulation models and data challenges*

10/2005 – Marcello D' Orazio, Marco Di Zio e Mauro Scanu – *A comparison among different estimators of regression parameters on statistically matched files through an extensive simulation study*

11/2005 – Stefania Macchia, Manuela Murgia, Loredana Mazza, Giorgia Simeoni, Francesca Di Patrizio, Valentino Parisi, Roberto Petrillo e Paola Ungaro – *Una soluzione per la rilevazione e codifica della Professione nelle indagini CATI*

12/2005 – Piero D. Falorsi, Monica Scannapieco, Antonia Boggia e Antonio Pavone – *Principi Guida per il Miglioramento della Qualità dei Dati Toponomastici nella Pubblica Amministrazione*

13/2005 – Ciro Baldi, Francesca Ceccato, Silvia Pacini e Donatella Tuzi – *La stima anticipata OROS sull'occupazione. Errori, problemi della metodologia attuale e proposte di miglioramento*

14/2005 – Stefano De Francisci, Giuseppe Sindoni e Leonardo Tininini – *Da Winci/MD: un sistema per data warehouse statistici sul Web*

15/2005 – Gerardo Gallo e Evelina Palazzi – *I cittadini italiani naturalizzati: l'analisi dei dati censuari del 2001, con un confronto tra immigrati di prima e seconda generazione*

16/2005 – Saverio Gazzelloni, Mario Albisinni, Lorenzo Bagatta, Claudio Ceccarelli, Luciana Quattrociocchi, Rita Ranaldi e Antonio Toma – *La nuova rilevazione sulle forze di lavoro: contenuti, metodologie, organizzazione*

17/2005 – Maria Carla Congia – *Il lavoro degli extracomunitari nelle imprese italiane e la regolarizzazione del 2002. Prime evidenze empiriche dai dati INPS*

18/2005 – Giovanni Bottazzi, Patrizia Cella, Giuseppe Garofalo, Paolo Misso, Mariano Porcu e Marianna Tosi – *Indagine pilota sulla nuova imprenditorialità nella Regione Sardegna. Relazione Conclusiva*

19/2005 – Fabrizio Martire e Donatella Zindato – *Le famiglie straniere: analisi dei dati censuari del 2001 sui cittadini stranieri residenti*

20/2005 – Ennio Fortunato – *Il Sistema di Indicatori Territoriali: percorso di progetto, prospettive di sviluppo e integrazione con i processi di produzione statistica*

21/2005 – Antonella Baldassarini e Danilo Birardi – *I conti economici trimestrali: un approccio alla stima dell'input di lavoro*

22/2005 – Francesco Rizzo, Dario Camol e Laura Vignola – *Uso di XML e WEB Services per l'integrazione di sistemi informativi statistici attraverso lo standard SDMX*

1/2006 – Ennio Fortunato – *L'analisi integrata delle esigenze informative dell'utenza Istat: Il contributo del Sistema di Indicatori Territoriali*

2/2006 – Francesco Altarocca – *I design pattern nella progettazione di software per il supporto alla statistica ufficiale*

3/2006 – Roberta Palmieri – *Le migranti straniere: una lettura di genere dei dati dell'osservatorio interistituzionale sull'immigrazione in provincia di Macerata*

4/2006 – Raffaella Amato, Silvia Bruzzone, Valentina Delmonte e Lidia Fagiolo – *Le statistiche sociali dell'ISTAT e il fenomeno degli incidenti stradali: un'esperienza di record linkage*

5/2006 – Alessandro La Rocca – *Fuzzy clustering: la logica, i metodi*

6/2006 – Raffaella Cascioli – *Integrazione dei dati micro dalla Rilevazione delle Forze di Lavoro e dagli archivi amministrativi INPS: risultati di una sperimentazione sui dati campione di 4 province*

7/2006 – Gianluca Brogi, Salvatore Cusimano, Giuseppina del Vicario, Giuseppe Garofalo e Orietta Patacchia – *La realizzazione di Asia Agricoltura tramite l'utilizzo di dati amministrativi: il contenuto delle fonti e i risultati del processo di integrazione*

8/2006 – Simonetta Cozzi – *La distribuzione commerciale in Italia: caratteristiche strutturali e tendenze evolutive*

9/2006 – Giovanni Seri – *A graphical framework to evaluate risk assessment and information loss at individual level*

10/2006 – Diego Bellisai, Annalisa Lucarelli, Maria Anna Pennucci e Fabio Rapiti – *Feasibility studies for the coverage of public institutions in sections N and O*

11/2006 – Diego Bellisai, Annalisa Lucarelli, Maria Anna Pennucci e Fabio Rapiti – *Quarterly labour cost index in public education*

12/2006 – Silvia Montagna, Patrizia Collesi, Florinda Damiani, Danila Fulgenzio, Maria Francesca Loporcaro e Giorgia Simeoni – *Nuove esperienze di rilevazione della Customer Satisfaction*

13/2006 – Lucia Coppola e Giovanni Seri – *Confidentiality aspects of household panel surveys: the case study of Italian sample from EU-SILC*

14/2006 – Lidia Brondi – *L'utilizzazione delle surveys per la stima del valore monetario del danno ambientale: il metodo della valutazione contingente*

15/2006 – Carlo Boselli – *Le piccole imprese leggere esportatrici e non esportatrici: differenze di struttura e di comportamento*

16/2006 – Carlo De Greogorio – *Il nuovo impianto della rilevazione centralizzata del prezzo dei medicinali con obbligo di prescrizione*

1/2007 – Paolo Roberti, Maria Grazia Calza, Filippo Oropallo e Stefania Rossetti – *Knowledge Databases to Support Policy Impact Analysis: the EuroKy-PIA Project*

2/2007 – Ciro Baldi, Diego Bellisai, Stefania Fivizzani, e Marina Sorrentino – *Production of job vacancy statistics: coverage*

Documenti ISTAT(*)

1/2002 – Paolo Consolini e Rita De Carli - *Le prestazioni sociali monetarie non pensionistiche: unità di analisi, fonti e rappresentazione statistica dei dati*

2/2002 – Stefania Macchia - *Sperimentazione, implementazione e gestione dell'ambiente di codifica automatica della classificazione delle Attività economiche*

3/2002 – Maria De Lucia - *Applicabilità della disciplina in materia di festività nel pubblico impiego*

4/2002 – Roberto Gismondi, Massimo Marciani e Mauro Giorgetti - *The italian contribution towards the implementantion of an european transport information system: main results of the MESUDEMO project*

5/2002 – Olimpio Cianfarani e Sauro Angeletti - *Misure di risultato e indicatori di processo: l'esperienza progettuale dell'Istat*

6/2002 – Riccardo Carbini e Valerio De Santis – *Programma statistico nazionale: specifiche e note metodologiche per la compilazione delle schede identificative dei progetti*

7/2002 – Maria De Lucia – *Il CCNL del personale dirigente dell'area 1 e la valutazione delle prestazioni dei dirigenti*

8/2002 – Giuseppe Garofalo e Enrica Morganti – *Gruppo di lavoro per la progettazione di un archivio statistico sui gruppi d'impresa*

1/2003 – Francesca Ceccato, Massimiliano Tancioni e Donatella Tuzi – *MODSIM-P:Il nuovo modello dinamico di previsione della spesa pensionistica*

2/2003 – Anna Pia Mirto – *Definizioni e classificazioni delle strutture ricettive nelle rilevazioni statistiche ufficiali sull'offerta turistica*

3/2003 – Simona Spirito – *Le prestazioni assistenziali monetarie non pensionistiche*

4/2003 – Maria De Lucia – *Approfondimenti di alcune tematiche inerenti la gestione del personale*

5/2003 – Rosalia Coniglio, Marialuisa Cugno, Maria Filmeno e Alberto Vitalini – *Mappatura della criminalità nel distretto di Milano*

6/2003 – Maria Letizia D'Autilia – *I provvedimenti di riforma della pubblica amministrazione per l'identificazione delle "Amministrazioni pubbliche" secondo il Sec95: analisi istituzionale e organizzativa per l'anno 2000*

7/2003 – Francesca Gallo, Pierpaolo Massoli, Sara Mastrovita, Roberto Merluzzi, Claudio Pauselli, Isabella Siciliani e Alessandra Sorrentino – *La procedura di controllo e correzione dei dati Panel Europeo sulle famiglie*

8/2003 – Cinzia Castagnaro, Martina Lo Conte, Stefania Macchia e Manuela Murgia – *Una soluzione in-house per le indagini CATI: il caso della Indagine Campionaria sulle Nascite*

9/2003 – Anna Pia Maria Mirto e Norina Salamone – *La classificazione delle strutture ricettive turistiche nella normativa delle regioni italiane*

10/2003 – Roberto Gismondi e Anna Pia Maria Mirto – *Le fonti statistiche per l'analisi della congiuntura turistica: il mosaico italiano*

11/2003 – Loredana Di Consiglio e Stefano Falorsi – *Alcuni aspetti metodologici relativi al disegno dell'indagine di copertura del Censimento Generale della Popolazione 2001*

12/2003 – Roberto Gismondi e Anna Rita Giorgi – *Struttura e dinamica evolutiva del comparto commerciale al dettaglio: le tendenze recenti e gli effetti della riforma "Bersani"*

13/2003 – Donatella Cangialosi e Rosario Milazzo – *Fabbisogni formativi degli Uffici comunali di statistica: indagine rapida in Sicilia*

14/2003 – Agostino Buratti e Giovanni Salzano – *Il sistema automatizzato integrato per la gestione delle rilevazioni dei documenti di bilancio degli enti locali*

1/2004 – Giovanna Brancato e Giorgia Simeoni – *Tesauri del Sistema Informativo di Documentazione delle Indagini (SIDI)*

2/2004 – Corrado Peperoni – *Indagine sui bilanci consuntivi degli Enti previdenziali: rilevazione, gestione e procedure di controllo dei dati*

3/2004 – Marzia Angelucci, Giovanna Brancato, Dario Camol, Alessio Cardacino, Sandra Maresca e Concetta Pellegrini – *Il sistema ASIMET per la gestione delle Note Metodologiche dell'Annuario Statistico Italiano*

4/2004 – Francesca Gallo, Sara Mastrovita, Isabella Siciliani e Giovanni Battista Arcieri – *Il processo di produzione dell'Indagine ECHP*

5/2004 – Natale Renato Fazio e Carmela Pascucci – *Gli operatori non identificati nelle statistiche del commercio con l'estero: metodologia di identificazione nelle spedizioni "groupage" e miglioramento nella qualità dei dati*

6/2004 – Diego Moretti e Claudia Rinaldelli – *Una valutazione dettagliata dell'errore campionario della spesa media mensile familiare*

7/2004 – Franco Mostacci – *Aspetti Teorico-pratici per la Costruzione di Indici dei Prezzi al Consumo*

8/2004 – Maria Frustaci – *Glossario economico-statistico multilingua*

9/2004 – Giovanni Seri e Maurizio Lucarelli – *"Il Laboratorio per l'analisi dei dati elementari (ADELE): monitoraggio dell'attività dal 1999 al 2004"*

10/2004 – Alessandra Nuccitelli, Francesco Bosio e Luciano Fioriti – *L'applicazione RECLINK per il record linkage: metodologia implementata e linee guida per la sua utilizzazione*

1/2005 – Francesco Cuccia, Simone De Angelis, Antonio Laureti Palma, Stefania Macchia, Simona Mastroluca e Domenico Perrone – *La codifica delle variabili testuali nel 14° Censimento Generale della Popolazione*

2/2005 – Marina Peci – *La statistica per i Comuni: sviluppo e prospettive del progetto Sisco.T (Servizio Informativo Statistico Comunale. Tavole)*

3/2005 – Massimiliano Renzetti e Annamaria Urbano – *Sistema Informativo sulla Giustizia: strumenti di gestione e manutenzione*

4/2005 – Marco Broccoli, Roberto Di Giuseppe e Daniela Pagliuca – *Progettazione di una procedura informatica generalizzata per la sperimentazione del metodo Microstrat di coordinamento della selezione delle imprese soggette a rilevazioni nella realtà Istat*

5/2005 – Mauro Albani e Francesca Pagliara – *La ristrutturazione della rilevazione Istat sulla criminalità minorile*

6/2005 – Francesco Altarocca e Gaetano Sberno – *Progettazione e sviluppo di un "Catalogo dei File Grezzi con meta-dati di base" (CFG) in tecnologia Web*

* ultimi cinque anni

7/2005 – Salvatore F. Allegra e Barbara Baldazzi – *Data editing and quality of daily diaries in the Italian Time Use Survey*

8/2005 – Alessandra Capobianchi – *Alcune esperienze in ambito internazionale per l'accesso ai dati elementari*

9/2005 – Francesco Rizzo, Laura Vignola, Dario Camol e Mauro Bianchi – *Il progetto "banca dati della diffusione congiunturale"*

10/2005 – Ennio Fortunato e Nadia Mignolli – *I sistemi informativi Istat per la diffusione via web*

11/2005 – Ennio Fortunato e Nadia Mignolli – *Sistemi di indicatori per l'attività di governo: l'offerta informativa dell'Istat*

12/2005 – Carlo De Gregorio e Stefania Fatello – *L'indice dei prezzi al consumo dei testi scolastici nel 2004*

13/2005 – Francesco Rizzo e Laura Vignola – *RSS: uno standard per diffondere informazioni*

14/2005 – Ciro Baldi, Diego Bellisai, Stefania Fivizzani, Annalisa Lucarelli e Marina Sorrentino – *Launching and implementing the job vacancy statistics*

15/2005 – Stefano De Francisci, Massimiliano Renzetti, Giuseppe Sindoni e Leonardo Tininini – *La modellazione dei processi nel Sistema Informativo Generalizzato di Diffusione dell'ISTAT*

16/2005 – Ennio Fortunato e Nadia Mignolli – *Verso il Sistema di Indicatori Territoriali: rilevazione e analisi della produzione Istat*

17/2005 – Raffaella Cianchetta e Daniela Pagliuca – *Soluzioni Open Source per il software generalizzato in Istat: il caso di PHPSurveyor*

18/2005 – Gianluca Giuliani e Barbara Boschetto – *Gli indicatori di qualità dell'Indagine continua sulle Forze di Lavoro dell'Istat*

19/2005 – Rossana Balestrino, Franco Garritano, Carlo Cipriano e Luciano Fanfoni – *Metodi e aspetti tecnologici di raccolta dei dati sulle imprese*

1/2006 – Roberta Roncati – *www.istat.it (versione 3.0) Il nuovo piano di navigazione*

2/2006 – Maura Seri e Annamaria Urbano – *Sistema Informativo Territoriale sulla Giustizia: la sezione sui confronti internazionali*

3/2006 – Giovanna Brancato, Riccardo Carbini e Concetta Pellegrini – *SIQual: il sistema informativo sulla qualità per gli utenti esterni*

4/2006 – Concetta Pellegrini – *Soluzioni tecnologiche a supporto dello sviluppo di sistemi informativi sulla qualità: l'esperienza SIDI*

5/2006 – Maurizio Lucarelli – *Una valutazione critica dei modelli di accesso remoto nella comunicazione di informazione statistica*

6/2006 – Natale Renato Fazio – *La ricostruzione storica delle statistiche del commercio con l'estero per gli anni 1970-1990*

7/2006 – Emilia D'Acunto – *L'evoluzione delle statistiche ufficiali sugli indici dei prezzi al consumo*

8/2006 – Ugo Guarnera, Orietta Luzi e Stefano Salvi – *Indagine struttura e produzioni delle aziende agricole: la nuova procedura di controllo e correzione automatica per le variabili su superfici aziendali e consistenza degli allevamenti*

9/2006 – Maurizio Lucarelli – *La regionalizzazione del Laboratorio ADELE: un'ipotesi di sistema distribuito per l'accesso ai dati elementari*

10/2006 – Alessandra Bugio, Claudia De Vitiis, Stefano Falorsi, Lidia Gargiulo, Emilio Gianicolo e Alessandro Pallara – *La stima di indicatori per domini sub-regionali con i dati dell'indagine: condizioni di salute e ricorso ai servizi sanitari*

11/2006 – Sonia Vittozzi, Paola Giacchè, Achille Zuchegna, Piero Crivelli, Patrizia Collesi, Valerio Tiberi, Alexia Sasso, Maurizio Bonsignori, Giuseppe Stassi e Giovanni A. Barbieri – *Progetto di articolazione della produzione editoriale in collane e settori*

12/2006 – Alessandra Coli, Francesca Tartamella, Giuseppe Sacco, Ivan Faiella, Marcello D'Orazio, Marco Di Zio, Mauro Scanu, Isabella Siciliani, Sara Colombini e Alessandra Masi – *La costruzione di un Archivio di microdati sulle famiglie italiane ottenuto integrando l'indagine ISTAT sui consumi delle famiglie italiane e l'Indagine Banca d'Italia sui bilanci delle famiglie italiane*

13/2006 – Ersilia Di Pietro – *Le statistiche del commercio estero dell'Istat: rilevazione Intrastat*

14/2006 – Ersilia Di Pietro – *Le statistiche del commercio estero dell'Istat: rilevazione Extrastat*

15/2006 – Ersilia Di Pietro – *Le statistiche del commercio estero dell'Istat: comparazione tra rilevazione Intrastat ed Extrastat*

16/2006 – Fabio M. Rapiti – *Short term statistics quality Reporting: the LCI National Quality Report 2004*

17/2006 – Giampiero Siesto, Franco Branchi, Cristina Casciano, Tiziana Di Francescantonio, Piero Demetrio Falorsi, Salvatore Filiberti, Gianfranco Marsigliesi, Umberto Sansone, Ennio Santi, Roberto Sanzo e Alessandro Zeli – *Valutazione delle possibilità di uso di dati fiscali a supporto della rilevazione PMI*

18/2006 – Mauro Albani – *La nuova procedura per il trattamento dei dati dell'indagine Istat sulla criminalità*

19/2006 – Alessandra Capobianchi – *Review dei sistemi di accesso remoto: schematizzazione e analisi comparativa*

20/2006 – Francesco Altarocca – *Gli strumenti informatici nella raccolta dei dati di indagini statistiche: il caso della Rilevazione sperimentale delle tecnologie informatiche e della comunicazione nelle Pubbliche Amministrazioni locali*

1/2007 – Giuseppe Stassi – *La politica editoriale dell'Istat nel periodo 1996-2004: collane, settori, modalità di diffusione*

2/2007 – Daniela Ichim – *Microdata anonymisation of the Community Innovation Survey data: a density based clustering approach for risk assessment*

3/2007 – Ugo Guarnera, Orietta Luzi e Irene Tommasi – *La nuova procedura di controllo e correzione degli errori e delle mancate risposte parziali nell'indagine sui Risultati Economici delle Aziende Agricole (REA)*