The system of short term business statistics on labour in Italy. The challenges of data integration

C. Baldi, D. Bellisai, F. Ceccato, S. Pacini,
L. Serbassi, M. Sorrentino, D. Tuzi

Istat DICS/DCSC/OCC, Rome, Via Tuscolana, n.1786
(baldi@istat.it, bellisai@istat.it, ceccato@istat.it, pacini@istat.it, laserbas@istat.it, mrsorren@istat.it, tuzi@istat.it)

Abstract:
Italy produces labour market short term statistics both for national releases and for EU (STS, LCI and JV) regulations through a system of three surveys: the census monthly survey on Large Enterprises (LES), the quarterly sample survey on job vacancies and hours worked (VELA), and the survey on employment, wages and labour cost (OROS) mainly based on social security data. This paper describes the rationale behind the integration of the three sources into a system and its maintenance over time. Record linkage is used to integrate administrative and survey data both for the definition of the current target population and for editing, imputation and grossing up. Aims of the system are to ensure timely data, consistency among series and over time.

Keywords: data integration, consistency

1. Introduction

Italy produces short-term business statistics on the labour market through a system of three sources: the census monthly survey on large enterprises (LES), the quarterly sample survey on job vacancies and hours worked (VELA), and a survey on employment, wages and labour cost (OROS) mainly based on social security data.
The system has been set up during the last decade under the pressure of the new EU regulations. At the beginning of 2000 Italy produced monthly data on employment, hours, wages and labour cost through the LES survey which covered, on a census basis, the population of firms with at least 500 employees. These nationally released indicators on the population of largest enterprises provided very timely information on the evolution of labour input while the task of producing more encompassing indicators was carried out by the National Accounts. The approval of the Short Term Statistics regulation (STS - 1998), the Labour Cost Index regulation (LCI - 2003) and the Job Vacancy statistics regulation (JV - 2008) has imposed an adaptation of the statistics production to comply with the standards required in terms of new target indicators (other labour costs, jobs vacancies) and population coverage (all firms with employees). Instead of a new large-scale sample survey, discouraged both by the considerable financial costs and the burden on enterprises, ISTAT started the OROS Project to exploit the social security (INPS) registers representing a low cost source of data on employment, wages and labour costs on the whole population of enterprises. Since INPS data did not contain any information on job vacancies and hours worked, except for paid time, a survey (VELA) was launched to collect information on these two variables. Due to budget constraints, this new survey was limited to the population of

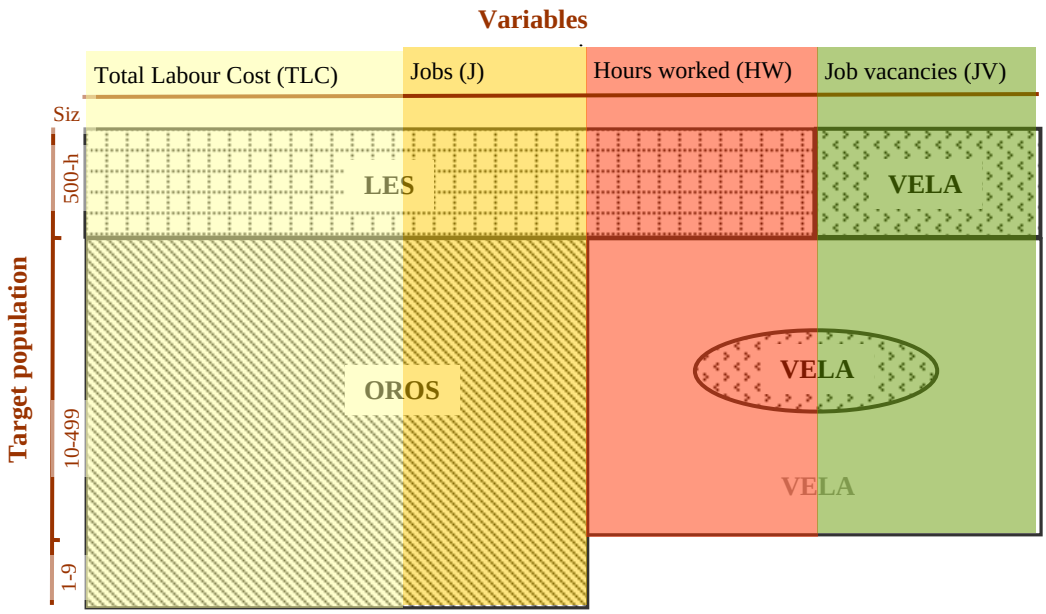
firms with at least 10 employees with take all strata for those over 500 employees. To limit the burden on enterprises the information on hours worked is requested only to the firms not responding to the LES survey, while, in order to check the data on vacancies, the questionnaire does not exempt these firms from providing information on jobs. In other words all the three sources contain information on jobs on the large enterprises (LEs); OROS and LES contain information on the labour cost variables on the LEs; OROS and VELA overlap for jobs on small and medium-sized enterprises (SMEs). All three sources cover firms belonging to the private business sector excluding agriculture (sections B to N of NACE Rev.2) and their economic activity code is mainly drawn from the statistical business register (ASIA).

The LES-OROS-VELA system has the objective of producing consistent quarterly estimates on jobs, wages and labour costs, hours worked and job vacancies on the population of firms with at least one employee, while continuing to produce the monthly survey figures for LEs. However, for the above mentioned constraints the statistics on job vacancies and hours worked are limited to the firms with at least 10 employees.

Figure 1 shows with respect to the three size classes subpopulations which sources are used for which variables. The pillars of the system are:

a) OROS+LES which is used both as the current quarter population frame and as the census based source of information for average quarterly jobs, wages and other labour costs;
b) VELA+LES which is used as the sample based source of information for jobs at the end of the quarter, job vacancies and hours worked.

Figure 1: *The integrated system: sources, variables and coverage.*



Although not coherent in all aspects, the system has some interesting features. It provides quarterly indicators on employment, wages and labour cost on a census basis. This census of a quarterly up-to-date population of all the firms with employees represents a substantial improvement for short-term statistics in Italy since the traditional sample surveys are all based on the Business Register. The delay of 15-24

months of the latter with respect to the current quarter implies difficulties in measuring the changes related to business demography. Another relevant aspect of the OROS-LES-VELA system is that the indicators produced are (internally) consistent for the overlapping firms with the LES monthly ones. Moreover, the quarterly indicators on job vacancies and hours worked are consistent with the estimate of jobs derivable from the OROS+LES subsystem on the population of firms with at least 10 employees. To preserve the consistency with the employment totals, the estimates of job vacancies and hours worked are obtained by reweighting the sample data to the portion of the current quarterly population with at least 10 employees.
One of the final outputs of this integrated system is the Labour Cost Index, described in paragraph 2, which is subject to strict timeliness constraints. Paragraph 3 illustrates the methodology used to build the subsystem OROS+LES with particular emphasis on the construction and maintenance over time of a unified list of enterprises. Paragraph 4 describes the procedures of micro integration needed to build the subsystem VELA+LES and the calibration to the OROS+LES universe. Some concluding remarks close the paper.

2. The LCI as an integrated output from the system

The LCI is a short-term indicator measuring the quarterly changes of the hourly labour cost and its single components (wages and salaries and other labour costs). Its transmission is due by 70 days from the end of the reference quarter and it is used by Eurostat to compile the aggregated Euro indicator on labour cost. The Italian NSI can nowadays satisfy the LCI regulation combining coherent and harmonized information on businesses produced by the LES-OROS-VELA integrated system on labour market statistics. Figure 2 gives a picture of the flow characterizing the system, from the inputs to the production of the main outputs, through the interrelation of three subsystems. Focusing on the LCI compilation, the flow chart illustrates how, starting from the data collection, the three surveys go through specific phases aimed at combining micro data and variables, up to the production of coherent indicators, both on labour cost and labour input variables.
To estimate the LCI, the hourly total labour cost indicator ($hwTLC^q$), with reference to quarter q, is derived as follows:

$$hwTLC^q = \frac{TLC^q}{THW^q} = \frac{jTLC^q_{OROS+LES} \cdot J^q_{OROS+LES}}{jTHW^{q^*}_{VELA+LES} \cdot J^q_{OROS+LES}} \qquad (1)$$

where $jTLC^q_{OROS+LES}$ is the per-capita indicator on total labour cost and $jTHW^q_{VELA+LES}$ is the per-capita indicator on hours actually worked. The three sources reconciliation is guaranteed by the number of jobs ($J^q_{OROS+LES}$) drawn from the OROS+LES subsystem and used by VELA as auxiliary variable for the estimation of hours worked (and job vacancies).

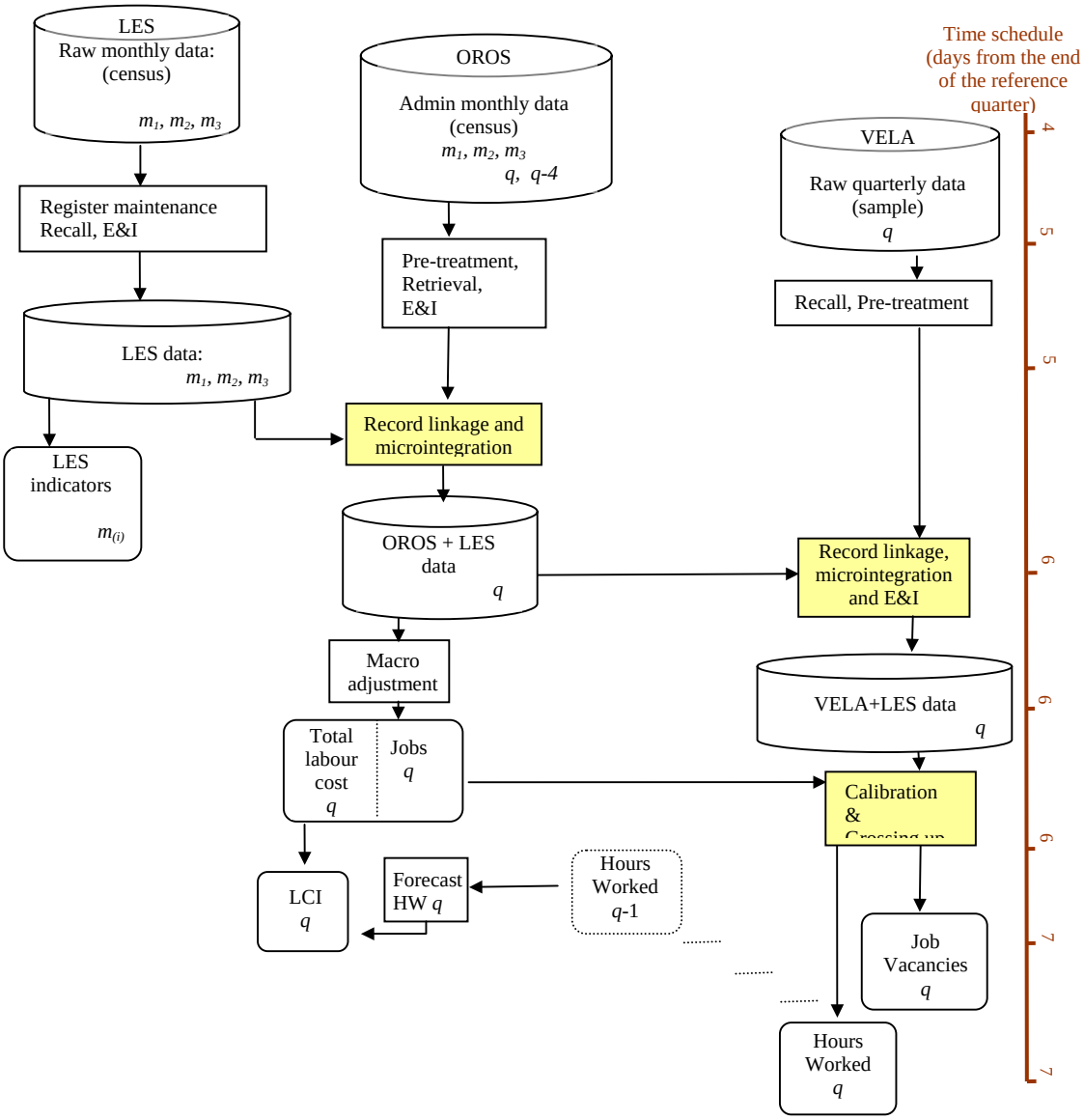
Dividing $hwTLC^q$ by the annual average of the same indicator calculated in the base year and applying the chain-linked Laspeyres formula, the LCI is finally obtained (Ciammola et al, 2009).
Figure 2 shows the three lines along which the whole production process is completed moving through more than one phase of integration: LES data *versus* OROS+LES data, OROS+LES data *versus* VELA+LES data. The scheme puts in evidence how reaching

the output depends on the strict keeping of the scheduled deadlines in a short-term context (right bar in figure 2).
At the beginning of the process, the validated monthly LES data are available at about 58 days from the end of the reference month. Once a year, with the release of the indicators referring to the first month, LES data for the whole previous year are revised in order to take into account late respondents and other updated information. LES data, for the three months of the quarter, are the first input for the system.
The OROS project, whose main goal is to cover the SMEs subpopulation, exploits administrative data. The source refers to the monthly contribution declarations that employers with at least one employee have to submit to INPS (Rapiti et al., 2010). Each quarter two main data sets are acquired: a set of preliminary data for quarter *q*, available after 45 days from the end of the quarter, that gives an almost complete coverage of the total OROS population, and is used for the preliminary estimates, and a final version of the data for quarter *q*-4. This second set is the basis to produce final estimates and differs from the preliminary version due to the late reporters and some measurement errors that have been corrected meanwhile by INPS. Given the administrative nature of the source, about 15 days are needed to make the data usable for statistical purposes, when a non-trivial preliminary phase of checks and computation is carried out[1] (Congia et al., 2008). At the same time, some structural information that contribute to the definition of the estimation domain (NACE code, etc.) is drawn from other statistical and administrative sources (Business Register, Tax Register etc.) and matched with OROS microdata using the fiscal code as unique linking key. After the calculation of the statistical variables, checks are implemented at different levels in order to identify possible anomalous values, both at monthly and quarterly frequencies. These checks are based on selective editing rules that exploit cross-sectional and longitudinal relations among the analyzed variables.
At 60 days from the end of quarter *q*, the OROS microdata are substituted for by the LES survey data for the overlapping firms. The availability of LES data at 58 days from the end of the quarter starts the record linkage and micro-integration process in order to single out from OROS the common sub-population of firms to avoid double counting. This process implies taking into account variables harmonization and linkage issues (two days) and produces the quarterly OROS+LES microdata (see §3). The almost complete coverage of the target population implies that the basis for estimation is a sum of the quarterly OROS+LES data. The estimates obtained turn out to be very accurate for the ratio variables (per-capita wages, labour cost and other labour costs). However the jobs level estimate has to be adjusted to correct for the incompleteness of the OROS data file due to the late reporters. The method currently used is a macro approach that exploits past revision errors and other longitudinal information. It is carried out on data aggregated at the two-digit level of the NACE. The estimates of the total number of jobs are usually available 60-62 days after the end of the reference quarter.
The VELA survey completes the short-term statistical system with the production of the job vacancy variable and the extension of the hours worked coverage. Each quarter the data collection ends around 55 days after the end of the reference quarter, but a subsequent recalling phase implies that late responses are accepted until 59-60 days after the end of the quarter. The first phase of the integration OROS+LES *versus*

[1] The correct exploitation and the translation of these data into statistical information entail coping with frequent changes in laws, regulations and other technical aspects regarding social security contribution. The necessary metadata are not made functionally available by INPS but have to be collected and regularly updated into an electronic format.

VELA+LES starts with the OROS+LES combined data availability at 60 days. These microdata are used both as auxiliary variables for editing and imputation of the SMEs jobs collected by VELA and as microdata source for the LEs sub-population estimates of jobs and hours worked (see §4). Both item and unit non-responses to the VELA survey are imputed on LEs while only item non-responses are imputed on SMEs. The second phase of the integration becomes possible when the macroadjusted estimates of jobs are provided by OROS+LES at 60-62 days. These estimates are used as known totals in the calibration procedure to obtain the weights for grossing up to the total population with more than 10 employees (see §4). After calibration and the calculation of the aggregates on the study domains, both job vacancies and hours worked estimates are validated through micro and macro checks. At this step the priority is given to the compilation of the job vacancy indicators that needs to be transmitted to Eurostat within 70 days from the end of the quarter.

Figure 2: *The integration flow*



In this integrated system, the timeliness requirements for LCI purposes face some difficulties. The hours worked from VELA+LES are in fact provided at about 75 days from *q* and a flash estimate is not available as it is for quarterly job vacancies (see §4). Nevertheless, since LES data on hours worked are available at 58 days, the current per-capita hours worked can be forecast, using the available time series information from the VELA+LES indicators (Ceccato et al., 2011). In order to use LES as a leading indicator in the forecast, a quarterly aggregation procedure on monthly data and a harmonization treatment are needed, requiring four days of work. The VELA+LES hours worked available at 75 days are used to revise the LCI in the next release, because of their non-negligible effects on the indicator.

For the sake of completeness, it is worth stressing the interrelations between the three subsystems as regards revisions and their effects on the LCI quality. The LES revisions produce a "domino effect" on the other two processes: they are included in the OROS+LES quarterly data in the first quarter release (at June) of the year (*y)*, and affect the estimates of the four quarters of the previous year (*y*-1). Furthermore, due to the

availability of a final version of the OROS microdata, an additional cause of revision affects each quarter the *q*-4 TLC estimates. On the other hand, VELA introduces yearly revisions referring to: the four quarters of the previous year (*y*-1), to acquire the LES revisions; the four quarters of the year before the previous one (*y*-2), to take into account the *q*-4 revisions of the jobs estimates. The final outcome on the LCI is a combined and more extended revision effect, driven by the single subsystems: the quarters of the current year are revised each quarter (both *via* annual weights, and because of the availability of the hours worked on *q*-1); once a year all the eight quarters of the previous two years are revised. The relative long time period concerned by revisions and the unavailability of the last observed data affect the quality of the provided LCI indicator. The next improvements of the system integration should start by tackling these two aspects.

3. The administrative and survey data integration

The integration of OROS administrative data with the monthly large firms survey defines first of all the current target population frame for the LES-OROS-VELA short term business statistics system.

This integration was a necessity until 2004, justified by the fact that large firms were underrepresented in the early reporters (transmitting the administrative data through an electronic form) used for the OROS preliminary estimates (Baldi et al., 2004). Afterwards, when data supplied by INPS were almost complete following a legislation change that obliged all the firms to declare the social security contributions electronically, the integration became mainly a choice. Since large firms have a relevant influence on the estimates (around 1,400 units in NACE sections B to N, accounting for about 20 per cent of all employee jobs), the survey data, collected and processed monthly by a group of specialized workers with continuous contacts with the enterprises, guarantee a higher quality of the information.

The integration followed a feasibility study which had found a good degree of comparability between the target variables of the two sources. Of course a harmonization is necessary in particular for jobs, which are measured with respect to different definitions (see Amato and Pacini, 2004a). In fact, while LES collects the number of jobs at the beginning and at the end of the month, OROS measures the quarterly average of jobs having at least a hour paid in the month.

Moreover, since the publication of monthly figures has been possible only through the LES survey, this type of integration would have opened the possibility to provide quarterly figures by size classes consistent with the monthly LES ones.

This choice implied that for the LES survey, whose main objective in the past was the release of aggregate figures, the production of high quality micro data and meta information became a new important aim rather than being just a by-product.

The integration procedure aims at replacing admin with survey data for the overlapping enterprises. The main operation consists in identifying and excluding from the OROS source the enterprises belonging to the LES survey.

In this process four specific features are noteworthy:

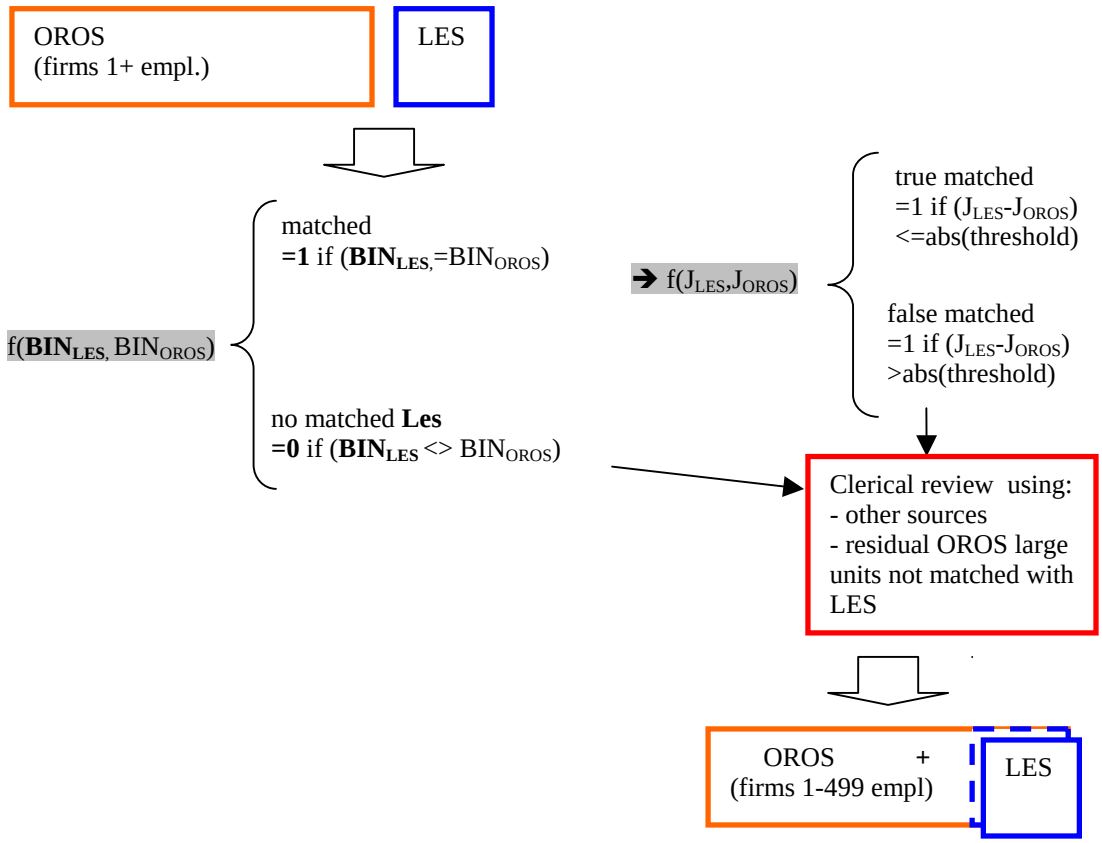
1. the linkage of the two sources is not a one-off operation but rather a process that must be carried out each quarter;
2. the dynamics of large firms due to corporate events must be monitored in order to guarantee the correct identification of the LES units in the OROS population;

3. to provide high quality data the acceptance threshold of mismatch errors has to be close to zero: the integration of the two sources must be carried out very carefully to avoid any misalignment and duplication;
4. considering the release calendars of LCI and JVS and the availability dates of OROS and LES data, the time allowed for the integration process is about 2 days (Figure 2).

To take into account all these aspects the integration process proceeds in two distinct steps. The first one, quite effort and time consuming, is carried out every five years when the census list of large enterprises is defined on the basis of the Business Register and administrative units in the OROS current population, with reference to the introduction of the new base year (STS, Reg. CE n.1165/1998). Starting from this LES list, a residual list of OROS units must then be identified. Theoretically speaking, in this year it would be possible to perform the integration procedure removing from OROS all the firms with more than 500 employees, and replacing them with LES data. In practice, a record linkage is needed because of the importance of defining a list to drag along beyond the base year in order to reduce the quarterly integration costs. This second phase consists in maintaining each quarter the complementarity of the two lists considering the LES list of the base year as a fixed panel. The panel definition implies that: no unit gets in or out of the lists as long as the base year remains the same, even if its size decreases crossing the 500 employees threshold; and all demographic company changes have to be considered to guarantee a longitudinal panel. Therefore, for example, new units resulting from a split up of a panel enterprise are included, as those deriving from a merger between panel and non panel firms.

Hence, the base activity to built the two complementary lists is performed in the base year (first step) mainly through an exact matching. Although the OROS and LES data have not the same statistical identifier, the former having an administrative code and the latter an internal survey code, the match is possible using the fiscal code as unique business identification number (BIN). To be used as key linking variable, this code must undergo a preprocessing treatment in both sources to guarantee a formally correct and never missing variable. Despite this accurate pre-matching process, some problems in using the BIN equality function as unique linkage key still remain, causing no matched and false matched pairs (Fig. 3). The first event occurs when two records belonging to the same unit are not linked, while false matches occur when the results of the function based on the BIN equality are positive but records belonging to substantially different units are linked (linkage key is not perfect and/or exhaustive). Two of the most important reasons explaining this phenomenon are the different rules for updating the register in the two sources together with the frequent occurrence of company changes which weaken the usefulness of the fiscal code as BIN. Information to follow the enterprises over time is needed but not available in the administrative data so their management is almost completely delegated to the LES survey experts. This information is stored in a data base of events where specific rules about registration and statistical treatment of business longitudinal changes are adopted taking into account the features of the LES panel (see Amato and Pacini, 2004b).

Figure 3: *The record linkage between LES and OROS data.*



In order to detect false matches, after the automatic linkage, another indicator function is considered, based on the difference in term of jobs between the two units with the same BIN. An acceptance threshold is established to take into account the slight residual difference in jobs after the harmonization of the LES ones. For specific sectors, characterized by a high turnover that implies wider differences in the two sources jobs definition (see §4), this threshold is higher. The matched units with an absolute jobs difference above the threshold and no matches are passed to a clerical review carried out using different sources of information such as the Statistical Business Register, some on line administrative sources or firms web sites[2]. To avoid any further mistake a test is carried out listing and checking if there are residual large firms in the OROS data not matched with LES ones.

The list of units, thus created, is updated quarterly in a very short time (second step). The linkage scheme described above is applied only to the panel firms which have undergone changes identifiable through both new BINs and new mismatches in the jobs function. This second step allows to rapidly keep a dynamic integration between the two sources. The efficiency and efficacy of the enhancement of the list is tightly related to the capability of LES survey to trace corporate/company changes, also in consideration of the fact that in the survey late respondents and missing data increase exactly for the units undergoing these changes.

[2] To find out the correct match also a record linkage based on other variables related to the identification of the units (such as firm name, addresses, telephone numbers, etc.) could be considered. At the moment this solution has not yet been implemented, due to the small weight of these residual units compared to the high cost of standardization of alphanumeric variables.

After the identification of the LES units in the OROS source, the economic variables administrative data are replaced with survey ones. The availability in the LES survey of detailed wages and other labour cost components allows the calculation of the OROS survey target variables.

4. Implications for editing and imputation and grossing up in an integrated system

The OROS+LES data are subsequently integrated with those collected by the VELA survey in order to produce indicators on hours worked and job vacancies. The integration occurs at two levels:
1) at the micro level, the information from LES and VELA is combined to produce a unified sample dataset with the fields on jobs, job vacancies and hours worked filled for every unit which has responded to VELA and for all those belonging to the LES panel. Furthermore, this unified sample dataset is linked at the micro level with the OROS+LES population;
2) at the macro level, the thus obtained sample is reweighted to represent the portion of the OROS+LES population with at least 10 employees.
The integrated sample dataset is obtained mainly through a deterministic record linkage based on fiscal code using the large enterprises list of the survey units created as shown in paragraph 3. The same technique is used also for the link with the OROS+LES population.
The features of the micro integration procedure are different depending whether the units belong to the LES survey or not.
For the first type of units the integration consists in acquiring the data on the number of jobs, job flows and hours worked, even for the units that have responded to VELA, from the LES survey (since they are carefully edited by expert personnel and imputed for all unit non responses), and add the job vacancy variable collected only by VELA. The acquisition poses no particular problems since the definitions of the variables are the same in the two surveys. In particular, for jobs at the beginning and at the end of the quarter the data collected by LES for the same dates are used. For hours worked during the quarter, the data collected (or imputed) by LES for the three months of the quarter are added up to a quarterly figure. For job vacancies, which are collected only by VELA, if a LES firm is a respondent in VELA and the difference between jobs as measured by the two surveys is limited, the figure thus collected on the vacancies to jobs ratio is multiplied by the LES end of quarter jobs to obtain the number of vacancies for that unit; otherwise, a hot-deck nearest neighbour donor imputation is carried out (see for example Chen and Shao, 2000).
For the units not belonging to the LES survey, the integration plays a role in the editing and imputation of the number of jobs collected by VELA through the comparison with the OROS variable but considering that the two jobs variables are measured with respect to different definitions (§3). In fact, while VELA collects the number of jobs at the beginning and at the end of the quarter, the figure available in the OROS quarterly data is the average over the three months of the quarter of the monthly number of employees with at least one hour of work paid in the month. These two measures show a very high correlation and the distribution of their differences is sharply concentrated around zero (see Bellisai, Pacini and Pennucci, 2005a and 2005b). The method used to check for large differences is a variant of the resistant fences one (see Thompson and

Sigman, 1999)[3], which is applied within classes defined by economic activity and turnover to take into account the differences between the two jobs variables. Beyond the differences caused by record linkage or measurement errors in one of the two sources[4], those related to definitions may occur particularly in firms with a high turnover of employees within the quarter. The construction of a turnover proxy is possible using the information collected by VELA on the number of hires and separations.

The ordered values of the score function allow to split the data into three subsets according to an empirically chosen threshold which takes into account a cost-benefit assessment:

- a critical flow, consisting of the records above the threshold, that will undergo an interactive check since they are classified as possibly contain influential errors;
- a non-critical flow, consisting of records between the fences and the threshold, whose errors can be non-influential. These observations will undergo a subsequent automatic check of internal consistency and possibly will be treated automatically;
- a flow of observations (those between the fences) reputed correct.

The observations of the non-critical flow which have not passed the subsequent check and those of the critical flow that in the interactive phase have been reputed incorrect are set to missing and passed to an imputation phase. Here a hot-deck nearest neighbour donor imputation is carried out where the matching variable on which the distance is computed is given by the OROS jobs (which is reasonable given their strong correlation with VELA ones). The imputation is performed within classes defined by economic activity and firm size.

The dataset resulting from the micro integration procedure is then grossed up to the population of units with at least 10 employees based on the OROS+LES data built as described in paragraph 3.

The grossing up is carried out through calibration, with jobs as measured in this reference population as auxiliary variable.

An exception to this rule is applied in the few cases of the firms (not in the LES panel) for which the difference between jobs as measured by VELA and OROS is above a threshold. In this case, to avoid the risk of unsuitably large or small grossing up weights, the number of jobs quarterly average as measured by VELA is used in place of OROS jobs as auxiliary variable.

The known totals of the auxiliary variable are based on the OROS+LES microdata, adjusted in a way that incorporates the macro correction, described in paragraph 2, adopted to overcome the problem of late reporters (see Ceccato et al., 2011).

The calibration procedure is performed within cells defined by economic activity and firm size. The initial weight is the inverse of the inclusion probability multiplied by the response rate for the units belonging to the non-LES portion and a unit weight for the firms belonging to the LES portion (these units have been drawn with certainty and subsequently all the unit non responses are imputed). The calibration weights for the

[3] In this case the score function is built using: the logarithm of the ratio between the OROS and VELA values, a magnitude adjustment exponent given by the maximum between VELA and OROS jobs and a lower (upper) fence given by the first (third) quartile minus (plus) a non-linear function of the interquartile range.

[4] Measurement errors in the OROS micro data are typically related to late reporters in one of the three months of the quarter or to the fact that some categories of employees (such as workers in non-agricultural firms who carry out a work which is classified as agricultural) are not included in the OROS administrative data.

LES units will in general be slightly different from one (generally larger) as the weights adjust for the non-responses of the non-LES large firms.
The editing and imputation and grossing up procedures described so far in this paragraph are used to produce the quarterly aggregate figures on hours worked and job vacancies on firms with at least 10 employees. However, the EU job vacancy regulation also requires the transmission of data at least on the B-N aggregate within 45 days after the end of the reference quarter. Because the LES data for the last month of the quarter and the OROS data for the quarter are not available in time to comply with this deadline, the above described procedures have been adapted for this purpose. In particular, the LES end of the second month of the quarter jobs are used as estimates of the LES end of quarter ones, and the OROS microdata on jobs on the previous quarter and the same quarter of the previous year (to take into account possible seasonal effects) are used in place of those for the reference quarter in editing and imputation and grossing up.

5. Concluding remarks

The Italian integrated system of administrative and survey data for the production of short term business statistics on the labour market has been implemented over the last 10 years to guarantee the compliance with EU regulations. There are of course unsolved problems and inefficiencies mainly linked to an integration process not designed before but realized ex post.
Some improvements to increase the efficiency and the quality of the general system are currently being studied/implemented. For example a project is under way to include the questions on job vacancies in the LES questionnaire starting from 2012, in order to reduce both the statistical burden on large firms (which would no more receive the VELA questionnaire) and the costs for the Italian NSI. It is also expected that this integration in the data collection phase would lead to a higher quality of the disseminated indicators due to the more relevant role of the LES experts and the smaller need for sample data micro integration. Moreover, in the OROS survey experimentations on the micro imputation of missing units (late reporters) are carried out, with the aim to substitute the macro adjustment. This would increase the quality of the microdata used in the editing and imputation of VELA jobs and in the reweighting procedures.
More complex would be to reach the timeliness in the production of hours worked indicators that would allow the use of the current quarter data on this variable in the LCI rather than their time series forecast. In fact, to achieve this goal, increases in the timeliness of each process are needed in a context where already resources are stretched and production times very tight. Also the fulfillment of the new European requests (such as, for example, the coverage of NACE Rev. 2 sections P to S) is a big challenge.
Furthermore, the implications for the system of the revision policies of the single sources should be more carefully considered.
Finally, the general Italian available statistical sources context and its evolution have to be considered. In recent months big changes are occurring for what concerns the availability of new administrative information with a good timeliness. The use of this new administrative sources for a yearly virtual business census could reduce the delay of some statistical register (such as, for example, the Business Register) and allow a

redesign of the sample surveys. This would imply a general reorganization of the system described here.

References

Amato G., Pacini S. (2004a), *L'Occupazione delle indagini GI e OROS a confronto*, mimeo, Istat.

Amato G., Pacini S. (2004b), *Le retribuzioni delle indagini GI e OROS a confronto*, mimeo, Istat.

Baldi C., Ceccato F., Cimino E., Congia M.C., Pacini S., Rapiti F., Tuzi D. (2004) Use of Administrative Data to produce Short Term Statistics on Employment, Wages and Labour Cost, in: *Essays*, n.15, Istat.

Bellisai D., Pacini S. e M.A. Pennucci (2005a), *Analisi preliminare sui dati Oros e Vela – 1a parte*, mimeo, Istat.

Bellisai D., Pacini S. e M.A. Pennucci (2005b), *Analisi preliminare sui dati Oros e Vela – 2a parte*, mimeo, Istat.

Ceccato F., Tuzi D. (2011) Labour Cost Index (LCI). Quality Report 2010. *Eurostat Internal Report*.

Ceccato F., Cimino E., Congia M.C., Pacini S., Rapiti F., Tuzi D. (2011) I nuovi indicatori trimestrali delle retribuzioni lorde, oneri sociali e costo del lavoro della rilevazione Oros in base 2005 e Ateco 2007, mimeo, Istat

Chen J., Shao J. (2000) Nearest neighbor imputation for survey data. *Journal of Official Statistics*, 16, 113–131.

Ciammola A., Ceccato F., Congia M. C., Pacini S., Rapiti F., Tuzi D. (2009) The Italian Labour Cost Index (LCI): sources and methods, *Contributi Istat*, vol.8, Rome (http://www3.istat.it/dati/pubbsci/contributi/Contributi/contr_2009/08_2009.pdf).

Congia M.C., Pacini S., Tuzi D. (2008) Quality Challenges in Processing Administrative Data to Produce Short-Term Labour Cost Statistics, in: *Proceedings of Q2008 European Conference on Quality in Official Statistics*, Rome (http://q2008.istat.it/sessions/paper/29Congia.pdf).

Rapiti F., Ceccato F., Congia M. C., Pacini S., Tuzi D. (2010) What have we learned in almost 10-years experience in dealing with administrative data for short term employment and wages indicators? *Paper presented at the ESSnet seminar Rome, March* (http://www.ine.pt/filme_inst/essnet/papers/Session2/Paper2.4.pdf).

Thompson, K.J., Sigman, R.S. (1999) Statistical Methods for Developing Ratio Edit Tolerances for Economic Data. *Journal of Official Statistics*, 15, 517-535