The micro-level integration of Labour Register and Compulsory Communications to provide timely and consistent stock-flows accounts of the labour market

Ciro Baldi[1], Sara Gigante[1], Silvia Pacini[1], Roberta Rizzi[1]

Abstract

An experimental integration at the micro level of each job position between an employer and an employee in the private non-agricultural sectors enables the extension of the stocks and flows of the Labour Register (LR) to include the much-updated flows of the Compulsory Communications (CC). This analysis makes it possible to correct the LR for the absence of very short-duration job positions and project job stocks and flows up to the last available date of the CC. Daily stocks and the relative activation and cessation flows allow detailed studies of the labour dynamic. The integration methodological details are examined in depth to highlight the main evidence and the points to enhance.

---

1   Ciro Baldi (baldi@istat.it); Sara Gigante (gigante@istat.it); Silvia Pacini (pacini@istat.it); Roberta Rizzi (rizzi@istat.it), Italian National Institute of Statistics – Istat.

---

1. Introduction

The current situation of administrative data and statistical registers in Italy allows the calculation of two sets of employment measures referred to jobs.

On one side there are stock-type statistics such as the number of active jobs on a specific day of the year and various kinds of averages over such figures over a given period (*i.e.* a month, a quarter, a year, *etc.*). These measures are by now produced by the Italian National Institute of Statistics - Istat trough the Oros and the Labour Register (LR)[2] processes based primarily on the Italian National Social Security Institute - INPS data. LR data are disseminated through the Statistical Business Register (BR) and its employment detailed version[3], the Labour Cost Survey, the National Accounts, *etc*.

On the other side, the Compulsory Communications (CC) administrative system allows the production of flow-type statistics such as the number of activated and ceased work relationships in any given period (a day, a month, a year, *etc*.). For instance, this type of statistics is what is actually disseminated by the Ministry of Labour in its quarterly publication or, after a certain data transformation[4], in the Quarterly Note and Annual Report conjointly produced by Istat, Ministry of Labour, INPS, *Agenzia Nazionale per le Politiche Attive del Lavoro* - ANPAL, Italian National Institute for Insurance against Accidents at Work – INAIL (aka "*Accordo a 5*"), and what has been monthly released, for almost 20 years now, by the *Veneto Lavoro* Research Centre.

The two sets of measures are, in theory, related to each other by the dynamic equation of job stocks and flows (see Section 6). This circumstance makes it possible to derive from each of them its own measures of gross changes, such as the difference in the job positions between two moments in time. However, since the actual statistics are based on independent sources, the estimated figures of change, even after having limited the comparison to a common target population, are often inconsistent. An obvious solution to this problem is to build a data system in which the different sources are integrated at the micro level and are thus able to provide at once consistent measures of job stocks and flows. But, up to now, there have been no attempts

---

2   The Labour Register is called (in Italian) *Registro Tematico del Lavoro* and its acronym is RTL.
3   The Business Register is called in Italy *Archivio Statistico Imprese Attive - ASIA* and its employment detailed version *ASIA Occupazione*.
4   See Methodological Note in the Quarterly Note of "*Accordo a 5*".

to build such a micro-based system. Two main obstacles have prevented this development. The first is the fact that the relevant administrative sources are held by different institutions. The second is that a number of methodological and data cleaning issues have to be faced and resolved to connect the two administrative sources.

However, in recent years, new conditions have arisen that are finally allowing this evolution of the labour statistics. The first is the ongoing Istat project of the Labour Register, a micro-level Employer/Employee Longitudinal Database with information on employment, hours, wages, social contributions, *etc*. The project of the LR entails that the main statistical unit of the Register is the job position, which is the association between an employer and an employee defined by a certain activation date. Among the various reasons to choose this statistical unit the possibility of linking the CC source was crucial since its data are implicitly based on the same unit.

The second circumstance was the birth of the "*Accordo a 5*", the coalition of the 5 above- mentioned institutions to produce statistics. This cooperation enables the transfer of macro and, more recently, microdata between the Ministry of Labour and Istat thus laying the foundations to extend the LR with the inclusion of CC information.

It is important to notice that it is already possible to produce consistent measures of job stocks and flows from the data of the Labour Register itself, without the integration of the CC source, for the very reason that the statistical unit of the LR is the job position and the activation and cessation dates are recorded in the Register. However, this operation has two main limitations. First, due to the nature of the input data source, the LR by itself is not able to measure the frequent intra-month job activations and cessations between a given employer and a given employee that are quite widespread in certain economic activities with the implication of overestimating the duration of some job positions. Second, and more important, the LR data is issued yearly with a time delay of about 8-10 months from the reference year with provisional data. The integration of the CC source should, in principle, provide the possibility of overcoming the two limitations thus allowing to produce the stocks and flows integrated statistics on a short-term basis.

2. The Labour Register's aims and principles

The Italian Labour Register is one of the components of the System of Integrated Registers that Istat has been implementing to face the challenges of the European Statistical System strategy for the current decade (*ESS Vision 2020*. ESSC, 2014).

The aims of improving the efficiency of the production processes and releasing more integrated and coherent indicators play a crucial role in the new system of registers. This is a system of micro-level statistical databases mainly derived from administrative sources, at the most detailed level of the information available, covering the whole population of statistical units (UNECE 2017).

In the context of labour statistics, the Labour Register aims to be the basis for a number of labour market indicators and analyses: from the construction of standard macro statistical indicators to the evaluation of labour policies and the release of longitudinal microdata standard files for the researchers (Baldi *et al.*, 2018).

With a complex and rich structure of information, the purpose of the LR is to cover all regular paid jobs active in the national territory in all sectors of economic activity, either private or public, dependent or independent.

The realisation of the LR implies that a number of different administrative sources have to be integrated. As for the population of employees, the primary sources on which it is based are social security data and, in a second order, tax data. Additional relevant sources, for the public sector, are payroll data and, for the population of self-employed workers, Chambers of commerce data and other sources. The employer-employee structure of the LR and its information at the maximum level of detail available in the input sources in terms of units, variables and time references, allow a modular use of the information. Depending on the target variable, it is possible to derive information referred to each day, week, month or year on three main statistical units: the individual, the economic unit, and the job position which is the specific unit of the LR.

In this way, the Register is linkable on the one hand to the Population Register, through the person ID, and on the other to the Economic Units Registers through the unit ID, with the obvious advantage of possibly matching a lot more information.

Moreover, the very specific unit of the LR is the job position, defined as the working relationship established between an economic unit and an individual with a starting date. It nables tracking the relationship between an employer and an employee since its inception and, over time, to construct the working career of each worker within and across employers and the status of employment. The LR, due to its longitudinal nature and statistical unit, allows measuring a very rich stock-flow accounting with figures on levels, gross and net changes of employment both in terms of workers and jobs. Moreover, each job position is characterised by a wealth of other attributes such as the type of contract, the working time, the occupational qualification, the workplace, *etc.* Other information available in the LR are gross wages, other labour costs, total labour cost, gross earnings, hours paid/worked, mainly with a monthly reference period.

In this analysis, we focus on the jobs stock-flows measurement for the subpopulation of employees in the non-agricultural sectors employed in private enterprises according to the Italian Business Register. This selection is mainly due to the advanced stage of development of the LR for this subpopulation[5]. In the context of this subpopulation, referred to dependent employment, the job position corresponds to the employment contract between an employer and an employee where the starting date is the date of activation of the contract and the ending date its cessation.

For the subpopulation we focus on, the main source of the LR is the UniEmens declaration of INPS which is integrated with the INPS-Dmag declaration for the agricultural workers employed in enterprises of the private non-agricultural sector. These two administrative declarations, made by employers for their employees with information on the job position and employer level, are used both for the social security payments and the individual pension schemes. Both these administrative sources have information to estimate either stocks or flows. To be more precise, the UniEmens dates of activation and cessation have to be integrated with the activation dates of tax

---

5   For the sake of precision the data for this subpopulation of the LR are derived from the RACLI Register. The RACLI Register is basically the frontline version of the module of the LR concerned with the population of employees with a job in enterprises of the private sectors covered by the BR. The standardised integration process of the sources and other production aspects like main metadata flowed into the LR process which is now an input sources of the RACLI process for the production of specific variables (worked and paid hours, labour costs, *etc.*) and their official release (see I.Stat database http://dati.istat.it/). Moreover the RACLI register is also the experimental area (for new variables, integration of new sources as CC, and so on) before introducing innovations into the LR.

sources for the job position activated before the first year of availability of the LR (2016). Moreover, in this social security source, when there are several job positions between the same employer and worker in the same month, only the earliest activation date and the latest closing date must be indicated in the declaration. This implies an undercounting of very short-term job positions within a month.

Regarding timeliness, both UniEmens and Dmag sources are available to Istat twice a year: one with provisional data and one with final data, both having a delay of 4 and 10 months from the end of the reference year respectively. Due to the timing of the production process and the availability of other sources, the first, provisional release of the LR data is available at about 10 months from the reference year and the final release at about 16 months.

In this work, the experimental analyses are carried out using the longitudinal LR information on job position stocks and flows related to the years from 2016 to 2019, with provisional data for this last year. As above mentioned from this data source alone it is possible to produce stocks and flows accounting of job positions. The analysis can be broken down across different characteristics of the job position, the employer and the employee. However, the analysis allowed by the LR alone has two main limitations. The first is the reference period: the timeliness of the administrative sources used at the moment in the LR process allows to produce preliminary estimation with a delay of about 10 months after the end of the reference year. This implies that estimates referred to January are released with a delay of 21 months. The second is related to the accuracy of the measurement of the flows of very short-term (intra-month) job positions for the reasons cited above. These limitations can be overcome by integrating into the LR the data of Compulsory Communication declarations.

3. The Compulsory Communications source: main features

The Compulsory Communications system, managed by the Ministry of labour and social policy, is a stream of declarations due by employers to notify relevant events of each employment relationship: the activation of a new employment relationship (contract) between an employer and a worker; the extension of a previous temporary contract; the transformation of the contract from one type to another (temporary to permanent, apprenticeship to permanent, part-time to full-time); and the termination of a contract before the natural deadline (not requested for contracts ending at the natural deadline) (Baldi *et al.*, 2014). The events contained in the declarations, per se, allow measuring only employment flows and not stocks of jobs or persons. However, the events pertaining to the same contract between an employer and an employee can be linked together to construct the job position. Unfortunately, since the CC system includes only the set of jobs which have been interested by an event since the start of the system (March 2008), all the permanent jobs created before 1 March 2008 and not subject, after that date, to termination, an extension or a transformation are not observed. Thus, the number of job positions active at any given date measured by the system is severely underestimated. While the system is unable to measure the level of the stocks it should in principle provide quite accurate estimates of their change. In fact, the net change in the number of jobs between two dates is simply equal to the difference between the number of activations and the number of cessations.

The CC declarations have information on the employer - identified through the tax code and characterised by economic activity sector, registered office, place of work, the worker - identified through the tax code and characterised by biographical variables such as gender, age, educational level, nationality, and the job (connecting employer and worker) - characterised by the date of activation, and eventually termination, established duration, type of contract, working time. As above mentioned, the job positions are built by the Ministry of Labour by sequentially linking all the declarations referred to the same relationship: one activation, (possibly) one transformation of fixed-term employment to permanent, (possibly) one or more extensions, and (possibly) one termination. The key, which permits the reconstruction of the job position and identifies it, is threefold: composed of the ID of the worker, the ID of

the employer and the starting date of the relationship. If one of the three key variables is affected by errors, a declaration is not matched with the correct job and the chain of related events is affected by errors.

The collection of this kind of data was initiated to provide an information system supporting actions to contrast irregular work but also to implement a database for monitoring and evaluating labour policies. Some features of this source make it valuable for carrying out a very innovative analysis of the labour market, especially when integrated with stock data offered for example by the Labour Register, as it is in the current work (see Section 5). The integration of the CC with the LR can contribute to improve the Register in a number of ways: full integration of LR with CC may allow to assess the quality of job definition, the most relevant statistical units of the Register, that have been defined through the threefold key (employer id, worker id, starting date) consistently with the CC source; checking and editing some variables already present in the Register with the corresponding variables available in CC; integrating variables absent in the Register (like professional qualification) or present with a lower level of detail (like the type of contract or place of work); improving metadata information thanks to further standard dictionaries and classifications. However for the current work the integration was limited to three main purposes (see Section 4): checking possible under coverage of the Register thanks to the very wide coverage in terms of employers and workers of the CC; identifying and measureing the labour input of the jobs with intramonth duration (those between the same employer and worker activated and terminated more than once in the same month) which cannot be identified by the LR *per se*; improving the timeliness of the information obtainable from the Register. This last point is due to the fact that since the CC declaration flows in the system daily, in principle they can be used to project the data of the LR, updated to the end of the previous year and provide preliminary, but very timely, estimation of the employment stocks with a very detailed breakdown.

4. Methodology

In order to improve the accuracy in estimating employment dynamics on LR basis, information declared in the administrative source of CC was integrated at the micro level of job position with the information estimated in the LR. Micro-integration is defined as "*the method that aims at improving the data quality in combined sources by searching and correcting for the errors on unit level*" (Bakker, 2011).

Integration at the level of individual employment (job position level) is also meant to obtain consistent and high-frequency stock and flow measures (number of active job positions every single day, for example), that enable to build indicators on the employment and job turnover as a synthesis of activation and cessation rates.

Another advantage in integrating the CC data with the Labour Register, in addition to the improvement in the estimation of the number of job positions and their duration, is producing very timely preliminary estimates of employment stocks and flows.

In terms of data quality, this exercise aims at improving *accuracy and timeliness* dimensions in LR (Eurostat, 2003; Statistics Canada, 2002; Rosén and Elvers, 1999), constrained to the maintenance of prior level quality in *consistency* (Wang and Strong, 1996; Batini and Scannapieco, 2006), *privacy and security and unique keys* (Daas et al., 2009) dimensions. According to the major experts in register-based statistics, quality represents indeed one of the main methodological issues to be tackled (Bakker and Daas, 2012).

The preliminary step to the exercise of micro-integration consisted in the harmonisation of reference periods, the completion of populations and the harmonisation of units between the two sources (Van der Laan, 2000).

In particular, the perimeter of the exercise is defined by all the job positions with an employee contract between a worker and a private enterprise whose economic activity belongs to NACE rev 2 sections B to S. In the exercise the temporary agencies and the job-on- call positions are excluded[6].

---

6 The exclusions of the temporary agencies depends on the unavailability, at the moment, of the CC information on temporary workers (that are collected via a dedicated form within the CC system), while the exclusions of the jobs on call depends on the fact that both in the LR and in CC these jobs positions have a large number of errors on the dates of cessations and on the fact that the duration of the job position for them is a poor measure of the labour input. However since these two segments of the labour force account for a relatively low share of total employment (respectively 1.2% and 2.7%) their exclusion does not affect the overall relevance of the work.

The population of enterprises is further refined by including only those units listed in the Statistical Business Register for at least one year from 2016 to 2019.

The operationalisation of the above perimeter has required to restrict both the LR and the CC source (completion of populations) on the set of enterprises listed in the BR (with the exception of temporary agencies) and on the employee contracts (with the exception of job-on-call) to be included. This last restriction has been based on the source-specific variables that characterise the employment contracts[7]. In LR the harmonisation of the unit has demanded, substantially, a simple operation of aggregation on the job position, the statistical unit derived from administrative data. In particular, the essential structure of both CC and longitudinal LR in a specific period can be described as a list of job positions between an employer and an employee active for at least one day in the period.

Table 4.1 - LR and CC basic structure - Theoretical scheme (a)

Employer	Employee	Job position	Activation date (AD)	Cessation date (CD)
A	John	A-John-1	A-John-1(AD)	A-John-1(CD)
…	…	…	…	…
A	John	A-John-N	A-John-N(AD)	
A	Sally	A-Sally-1	A- Sally-1(AD)	A- Sally-1 (CD)
…	…	…	…	…
H	John	H-John-1	H-John-1(AD)	H-John-1(CD)
…	…	…	…	…
Z	Jack	Z-Jack-1	Z-Jack-1(AD)	

(a) The data used are anonymised so they have only an ID number for employers and employees. Here proper nouns are used only for immediate comprehension.

The harmonisation of the reference period in the two sources has entailed the building of an *ad hoc* LR process for provisional data on 2019. The estimate of quantitative and qualitative variables to be attributed to each job and the harmonisation of other variables and classifications is postponed to a future extension of this work.

---

7   The selection in CC of positions with an employee type contract (excluding job-on-call workers) was carried out on the basis of variables that may differ from those present in the social security sources underlying the LR.

4.1 Longitudinal editing and imputation process and integration methodology

Since the administrative sources at the basis of the LR are not intended to record the job positions as above defined, the quality of the Activation (AD) and Cessation (CD) dates is not always optimal. Thus, the statistical unit needs to be derived (Wallgren and Wallgren, 2014; Daas and Ossen, 2011) with a process aimed at:

1. estimating the activation and cessation dates selecting only events compliant with the statistical definition (in line with what is stated in CC);

2. making dates consistent in the year and among years, in order to ensure a longitudinal coherence.

An iterative E&I process has been set to improve the quality of the LR unit estimation (about 2.5% of the positions are subject to correction and the incidence of jobs with no longitudinal inconsistencies post-E&I is 99.99%): in particular, for each employer-employee pair the process aims at making the dates consistent in the year and between years on the basis of the edit rules, defined by domain experts and summarised in Table 4.2.

Table 4.2 - LR rules for E&I on longitudinal coherence - Theoretical scheme

Edit rule number	Type of ER	Description
1		AD not null if CD previous not null
2	consistency	CD not null if AD next not null
3		AD <=CD
4	non-overlapping	AD >= CD_previous or CD<= AD_ next
5	non-duplication	AD = CD if AD_ next = AD
General		CD_previous <= AD <= CD <= AD_next

The editing and imputation process, while guaranteeing excellent results in terms of reducing logical violations in the longitudinal reading of the positions (Table 4.4), does not deal with the poor accuracy in measurement and detection for very short-term relationships due to:

1. Quality of UniEmens flow: in the presence of several job positions between the same employer and employee in a month, the earliest activation date and the most recent cessation date of the month must be indicated.

2. In LR's estimation of job positions the temporary interruption in sending administrative declarations does not automatically close the position. This implies that the register might record as one long-lasting job position, which in reality is a series of job positions for which no cessation date and subsequent activation dates were communicated.

These cases where the LR pair has a single long-lasting position instead of multiple positions, possibly interspersed with interruptions, represent the lack of accuracy in Register we are aiming to correct by integrating CC source. The desired result of integration is to identify all the active positions in the domain of interest in the period and attribute the most accurate duration to them. A particularly problematic aspect of this exercise is that the variables in the link key and the variables that are intended to be corrected through micro-integration are partially overlapped.

The criteria underlying micro-integration are based on simplicity (it is assumed that the sources are essentially not subject to errors, with the sole exception of the well-known LR underestimation of short-duration jobs) and standardisation (positions relating to the population were treated according to the same integration rules).

In particular, it is assumed that:

1. A job position between a given employer and a given employee is considered to be the same in the two sources if it presents in both the same activation date or the same cessation date (definition of link key).

2. In the event of a conflict of either cessation or activation dates for job positions linked in the two sources (on AD or CD) the one closest to the linked date is chosen (in this way it is intended to correct the lack of information regarding very short jobs in LR). Job positions linked on CD with different ADs are treated as one job position in the integrated register with the common CD and as AD the one which is closest to the CD. Job positions linked on AD with different CDs are treated as one job position in the integrated register with the common AD and as CD the one which is closest to the AD[8].

---

8 An additional validation of the duration of the jobs could also take into account the consistency with other register variables on labour input. This might furtherly reduce the risk of errors in the estimation of job duration, and particularly of underestimating the length of employment.

3. Job positions may exist and may not be present in LR: a job position between a given employer and a given employee found in the CC source for which no job position between the same employer and employee with the same AD and/or CD is found in the LR is treated as a job position to be added to the integrated register (this rule is consistent to employment projection on CC bases only as described in Section 5).

4. Job positions may exist and may not be declared in source CC: a job position between a given employer and a given employee found in the LR source for which no job position between the same employer and employee with the same AD and/or CD is found in the CC is treated as a job position to be added to the integrated register (this rule intend to correct the under coverage in the CC source that lacks all the job positions activated before the birth of the system - March 2008 - that had no change afterwards).

Table 4.3 - Theoretical scheme of LR and CC integration methodology to develop the Extended LR (E-LR)

A: Job positions with the same activation and cessation date	
Advantage	None
Risks	None
LR jobs	A- Sally – 1 $_{LR}$
CC jobs	A- Sally – 1 $_{CC}$
Linkage	A – Sally – 1 $_{LR}$ Link (on AD/CD) A- Sally – 1 $_{CC}$
AD estimate	A-Sally -1 (AD) $_{ELR}$ = A- Sally -1 (AD) $_{LR}$ = A- Sally -1 (AD) $_{CC}$
CD estimate	A- Sally -1 (CD) $_{ELR}$ = A- Sally -1 (CD) $_{LR}$ = A- Sally -1 (CD) $_{CC}$
E-LR jobs	A- Sally – 1 $_{ELR}$

B: Multiple short-term job positions not correctly identified in LR		
Advantage	Better accuracy in estimation	
Risks	None	
LR jobs	A- Sally – 1 $_{LR}$	
CC jobs	A- Sally – 1 $_{CC}$	A- Sally – 2 $_{CC}$
Linkage	A – Sally – 1 $_{LR}$ Link (on AD) A- Sally – 1 $_{CC}$	A – Sally – 1 $_{LR}$ Link (on CD) A- Sally – 2 $_{CC}$
AD estimate	A- Sally -1 (AD) $_{ELR}$ = A- Sally -1 (AD) $_{CC}$ = A- Sally -1 (AD) $_{LR}$	A- Sally -2 (AD) $_{ELR}$ = Max (A- Sally -1 (AD) $_{LR}$, A- Sally -2 (AD) $_{CC}$) = A- Sally -2 (AD) $_{CC}$
CD estimate	A- Sally -1 (CD) $_{ELR}$ = Min (A- Sally -1 (CD) $_{LR}$, A- Sally -1 (CD) $_{CC}$) = A- Sally -1 (CD) $_{CC}$	A- Sally -2 (CD) $_{ELR}$ = A- Sally -2 (CD) $_{CC}$ = A- Sally -1 (CD) $_{LR}$
E-LR jobs	A- Sally – 1 $_{ELR}$	A- Sally – 2 $_{ELR}$

Table 4.3 cont. - Theoretical scheme of LR and CC integration methodology to develop the Extended LR (E-LR)

	C: A long lasting job position not declared in CC		
Advantage	Accuracy in estimation of occupational stocks		
Risks	None		
LR jobs	A- Sally – 1 $_{LR}$		
CC jobs			
Linkage	No link		
AD estimate	A- Sally -1 (AD) $_{ELR}$ = A- Sally -1 (AD) $_{LR}$		
CD estimate	A- Sally -1 (CD) $_{ELR}$ = A- Sally -1 (CD) $_{LR}$		
E-LR jobs	A- Sally – 1 $_{ELR}$		
	D: A short-term job position not present in LR		
Advantage	Accuracy in estimation of occupational stocks		
Risks	Incorrect classification of an out-of-domain position in CC (over coverage()		
LR jobs			
CC jobs			A- Sally – 1 $_{CC}$
Linkage	No link		
AD estimate	A- Sally -1 (AD) $_{ELR}$ = A- Sally -1 (AD) $_{CC}$		
CD estimate	A- Sally -1 (CD) $_{ELR}$ = A- Sally -1 (CD) $_{CC}$		
E-LR jobs			A- Sally – 1 $_{ELR}$
	E: Job positions of different duration with the same activation date or cessation date		
Advantage	Correction of approximation in the estimation of a date after editing and imputation process in LR		
Risks	Underestimation of the length of employment		
LR jobs			A- Sally – 1 $_{LR}$
CC jobs			A- Sally – 1 $_{CC}$
Linkage	A – Sally – 1 $_{RLR}$ Link (on CD) A- Sally – 1 $_{CC}$		
AD estimate	A- Sally -1 (AD) $_{ELR}$ = Max (A- Sally -1 (AD) $_{LR}$, A- Sally -1 (AD) $_{Cc}$) = A- Sally -1 (AD) $_{CC}$		
CD estimate	A- Sally -1 (CD) $_{ELR}$ = A- Sally -1 (CD) $_{LR}$ = A- Sally -1 (CD) $_{CC}$		
E-LR jobs			A- Sally – 1 $_{ELR}$
	F: Job positions with different activation and cessation date (maybe partially overlapping)		
Advantage	Accuracy in estimation of occupational stocks and flows (when jobs are not partially overlapping)		
Risks	Duplication		
LR jobs	A- Sally – 1 $_{LR}$		
CC jobs		A- Sally – 1 $_{CC}$	A- Sally – 2 $_{CC}$
Linkage	No link	No link	No link
AD estimate	A- Sally -1 (AD) $_{ELR}$ = A- Sally -1 (AD) $_{LR}$	A- Sally -2 (AD) $_{ELR}$ = A- Sally -1 (AD) $_{CC}$	A- Sally -3 (AD) $_{ELR}$ = A- Sally -2 (AD) $_{CC}$
CD estimate	A- Sally -1 (CD) $_{ELR}$ = A- Sally -1 (CD) $_{LR}$	A- Sally -2 (CD) $_{ELR}$ = A- Sally -1 (CD) $_{CC}$	A- Sally -3 (CD) $_{ELR}$ = A- Sally -2 (CD) $_{CC}$
E-LR jobs	A- Sally – 1 $_{ELR}$	A- Sally – 2 $_{ELR}$	A- Sally – 3 $_{ELR}$

Table 4.3 presents a diagram of the possible cases that may arise when integrating LR and CC sources to derive an Extended LR (E-LR). Using the notation of Table 4.1 (employer-employee-job source), the possible cases in the relationship between employer A and employee Sally (see note (a) of Table 4.1) as declared in the sources are examined and the applied integration rule is listed highlighting its advantages and risks. In paragraph 4.2, all risks due to linkage errors (Fellegi and Sunter, 1969; Arts, Bakker and van Lith, 2000) or other micro-integration errors, and their impact on the integrated basis will be examined.

Another critical issue regards companies involved in legal changes and their employees. In order to simplify the methodology treatment, it was decided not to go into the matter in depth in this context. In this exercise, an attempt was made to minimise the impact of the differences in the enterprise's declarations in administrative sources by aligning the dates with the events reported by BR for transferred workers: in particular, for each event, it was necessary to align in both sources the cessation date within the transferring company and the activation date within the acquiring company[9].

4.2 Evaluation of methodological choices

With regard to the quality of the estimates in LR for the 2016-2019 ad hoc longitudinal data and the E-LR longitudinal data, some statistics on longitudinal consistency of positions are given below (Table 4.4).

The E&I process described above and applied on the LR should be performed also on the E-LR, which is after the micro-integration of the two sources. However, this was not done in this experimental work. This implies that the number of violations of the edit rules on the dates is higher in the E-LR than in the LR. Nonetheless, the incidence of violations remains very low: even summed up across rules this incidence amounts to 0.8% of the total number of job positions (Table 4.4), one-third of the incidence of positions with violations found in LR pre-E&I. In a similar way, the number of partial overlaps that are introduced by the integration (case F in Table 4.3) is very

---

9   In E-LR this justifies the possible change of dates for about 550 thousands of pairs (1.63% of total pairs) in case there is no consistency with the figure estimated by BR. In essence, the comparison leads to change in about 150 thousands of pairs (0.49% of the total) and finds alignment for the remaining 445 thousands (1.14%).

limited as well. This confirms the substantial consistency between the sources of social security, underlying the LR, and the CC source.

Table 4.4 - Number of jobs with E&I rules' violations in LR and in E - LR data - Years 2016-2019

Violated rule	LR		E-LR	
	N	%	N	%
Rule 1 (consistency)	267	0.001	118,841	0.349
Rule 2 (consistency)	1,756	0.005	0	0
Rule 3 (consistency)	0	0	14,490	0.043
Rule 4 (non-overlapping)	143	0	177,639	0.521
Rule 5 (non-duplication)	0	0	0	0
None	34,067,197	99.99	38,816,768	99.21
Total	34,069,363	100	39,127,738	100

Source: LR and extended LR to CC

Regarding the employer-employee pairs, in the considered period, the number of pairs found only in the CC source amount to 3% of the total number of pairs in the E-LR database (case D in Table 4.3). By construction of the database, these are workers present in the CC source as employees of co-present companies only. The lack of employment signals in the LR could be explained by the absence of contributory coverage for them in the period or by a classification of the type of contract not aligned between the one declared in CC with that derived in the LR on the basis of social security sources. Further study on this issue, especially through comparison with the tax data, may indicate the strategy to follow for them. On the opposite side, 22.53% of total pairs are not found in CC (case C in Table 4.3). These are likely the job positions activated before the birth of the CC system, which have not undergone changes afterwards, and thus are unobserved by the CC source. The evidence that the average ratio between the number of job positions and the number of employer/employee pairs is very close to 1 (1.06) might be an indication that supports the hypothesis.

Table 4.5 - Number of employer-employee pairs and number of job positions classified by the presence of employer-employee pairs in the sources - Years 2016-2019

	N° pairs	Incidence % pairs	N° job positions
Pairs with at least one job in both sources	21,464,468	74.29	30,792,010
Pairs with LR only jobs	6,510,693	22.53	6,894,091
Pairs with CC only jobs	917,459	3.18	1,441,637
Total	28,892,620	100.00	39,127,738

Source: Extended LR to CC

Looking at the outcome of integration among the 74.3% of couples present in both sources (Table 4.6), they developed about 31 million job positions: 72% of them have both cessation and activation dates coinciding in the two sources (case A in Table 4.3), 11.0% of the positions have the same activation date, but different cessation date, while 3.3% have the same cessation date, but different activation date. The total number of positions linked on activation or cessation date between the two sources is 86.3%. By their nature and construction, greater coverage and accuracy in the measurement of stocks in the LR and better quality in the measurement of flows in the CC source are expected. Positions activated/closed in the period (eventful positions), which are not in CC, represent 1.21% of positions for co-present pairs, while stable positions located only in source CC are 0.42%.

Table 4.6 - Outcome of job position's integration for employer–employee pairs present in both sources. Years 2016-2019

|	No		% job positions	
	Which of the tota	Total	Which of the total	Total
Same activation and cessation date in LR and CC				
Total		22,162,397		71.97
Same activation and different cessation date in LR and CC				
Different cessation date: CC choice	2,841,828		9.23	
Different cessation date: LR choice	555,851		1.81	
Total		3,397,679		11.04
Same cessation and different activation date in LR and CC				
Different activation date: CC choice	830,980		2.70	
Different activation date: LR choice	188,381		0.61	
Total		1,019,361		3.31

Table 4.6 cont. - Outcome of job position's integration for employer–employee pairs present in both sources. Years 2016-2019

	No		% job positions	
	Which of the tota	Total	Which of the total	Total
Different activation and cessation date in LR and CC				
LR only				
Stable positions in the period			0.62	
Eventful positions in the period			1.21	
Total		563,698		1.83
CC only				
Stable positions in the period			0.42	
Eventful positions in the period			11.43	
Total		3,648,875		11.85
Total		30,792,010		100.00

Source: Extended LR to CC

Looking now at the outcome of integration for all pairs[10] (Figure 4.1), the total number of positions linked between the two sources is the 67.93% on a total of 39,127,738 job position.

Figure 4.1 - Percentage composition of job positions in E-LR by source and economic activity of the employer. Years 2016-2019 (a)
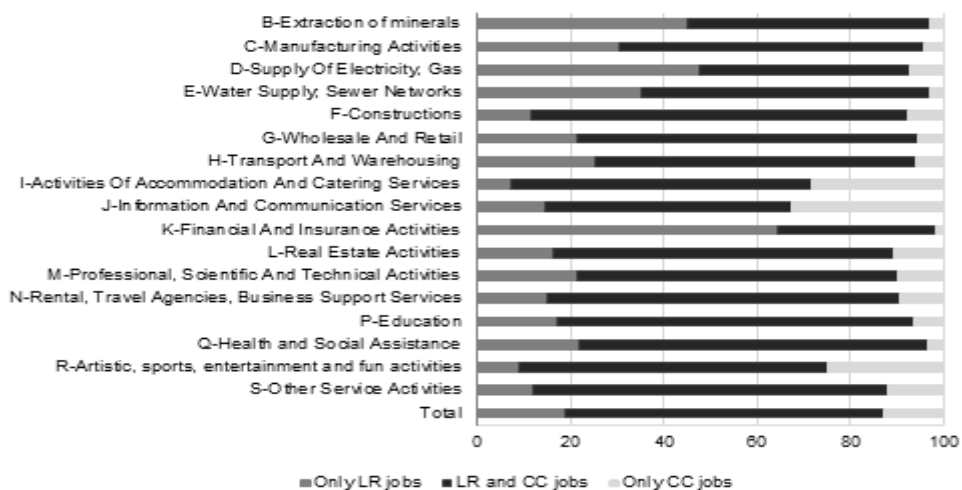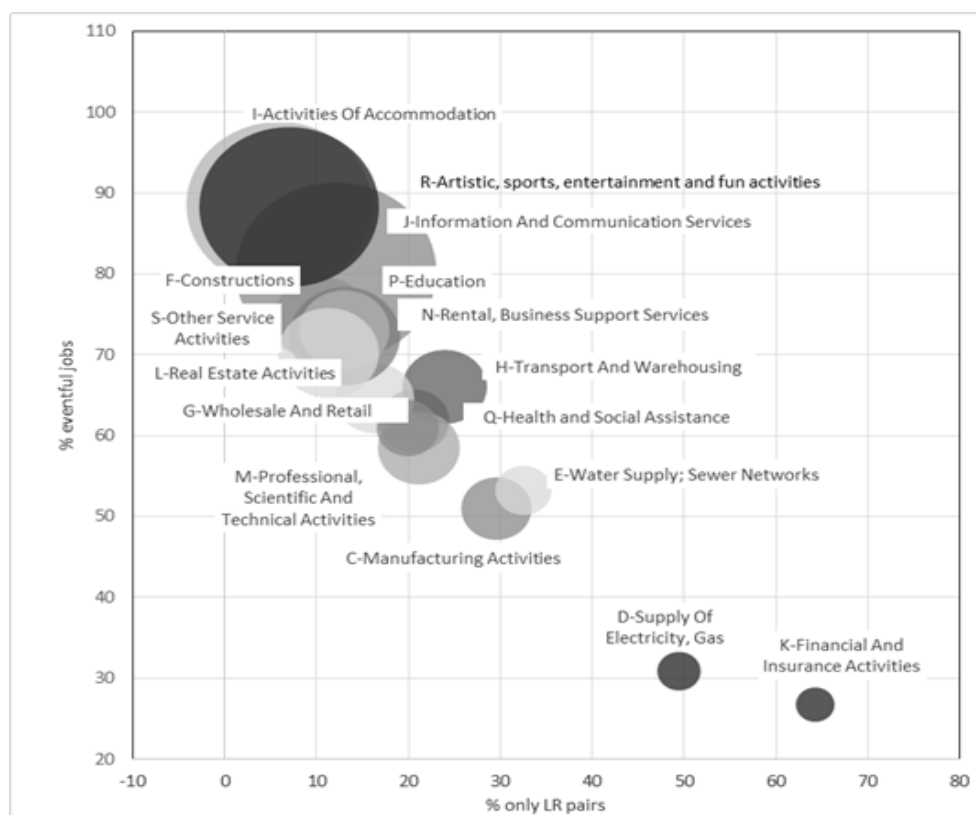

Source: Extended LR to CC

(a) The economic activity of the company is obtained using that attributed by BR for the most recent year available.

---

10 Correction made by merging positions that close on the same date even with different activation date is meant to correct also some overlapping position in CC source (0.3%)

The following figures show the correlation between the incidence of eventful positions, the incidence of only CC positions and the incidence of only LR pairs (Figure 4.2) and the correlation between the average duration of positions per pair and the incidence of only CC jobs. Figure 4.2 shows that as the incidence of eventful jobs increases the presence of only LR employer-employee pairs diminishes and the incidence of CC jobs only (size of bubbles) increases[11].

Figure 4.2 - Percentage incidence of eventful job position (Y value), only LR pairs (X value) and only CC job positions (circle area) by economic activity of employer. Years 2016-2019 (a)
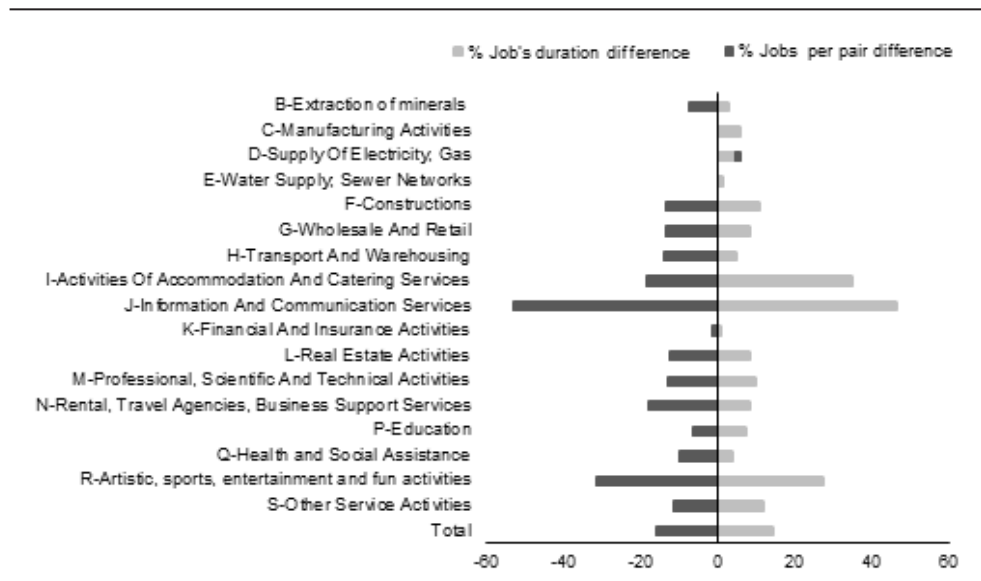

Source: Extended LR to CC
(a) The economic activity of the company is obtained using that attributed by BR for the most recent year available for observations in the LR and that declared in the CC source for the others.

11 From the figure are excluded some employer-employee pairs not present in the LR that declare in CC employer's activities not in the LR target population (Agriculture, Services to the person, Public Sector), underlining the problems of different classification by source due to misalignments in administrative declarations.

Figure 4.3 shows that the differences between LR and E-LR in the number of average positions per employer-employee pair and in the median duration of employment relationships are symmetrical. This evidence shows how the reduction in the duration of positions after integration with the CC source (16% in median) is due to a parallel increase in the number of relationships per pair (15%). Hence, the integration's aim to improve the accuracy in measuring short-term position seems to be accomplished.

Figure 4.3 - Percentage differences between LR and extended LR-CC in the number of average positions per employer-employee pair and in the median duration of job relationships. Years 2016-2019 (a)

Source: Extended LR to CC
(a) The economic activity of the company is obtained using that attributed by BR..

A joint reading of the two graphs highlights that, compared to LR, the percentage increase in the number of job positions per pair due to the integration is equal on average to 16% with the sectors with particularly high job turnover that record the largest increases. Sector R, for example, has an incidence of eventful jobs equal to 88% (Figure 4.2) and an increase in the number of job positions per pair of about 30% (Figure 4.3) aside from an incidence of only CC jobs of about 30%. The CC source, on the other hand, underreports the active positions in the low turnover sectors in the period (Supply of Electricity, Financial and Insurance Activities).

Below some evidences concerning the impact on integrated and corrected job positions (11.68% of total positions) of the methodological choices underlying integration are described. In particular, it is aimed at evaluating the possibility that the choice, in job positions characterised by the same pair and activation date, of the cessation date closest to the date of activation (*point a*)[12] or the choice, in job positions characterised by the same pair and cessation date, of the activation date closest to the cessation date (*point b*)[13] may underestimate the duration of the relationships (case E in Table 4.3). In order to assess the possible underestimate, the cessation date (activation date) in the source not chosen was compared with the maximum cessation date (minimum activation date) in the integrated base for the employer-employee pair in order to check whether several short positions in the integrated basis corresponded to a single position in the source.

Basically, the methodological choices seem to correct errors present in one of the two sources in most cases, although, of course, possible errors remain for a limited number of positions (about 183,612 positions corresponding to 0.47% of total positions and 4.02% of corrected positions). In general, the choice in LR to impute the cessation date (or activation date) only in the presence of a strong administrative signal (declaration of an activation date for the pair in the following months to impute the cessation date and the presence of a closing declaration to impute an activation date in the following months) has simplified and improved the integration of information from the most suitable source for making this attribution (CC source). Net of the few unresolved cases in the presence of inconsistencies between the sources the integration indeed ensured:

---

12 Considering *point a*, first we observe the positions for which the cessation date declared in CC prevailed (7.26%). Deepening each case, integration choices seem to be improving: remains only 1.7% (49,164 positions) for which the chosen method could underestimate the duration of the position of 45 days in median. Of them, 70% are fixed-term and, looking at some specific cases, there is a seemingly unjustified conflict between what is stated in UniEmens and what is present in CC. On the other hand, looking at the job positions for which the LR cessation date is chosen (1.42%), the possible approximation within the month for the dates imputed in LR could lead to a small error in the estimate (on average of 1 day, since however the approximation in LR concerns the days of the first week with contribution coverage in UniEmens) for 19% of the positions.

13 In deepening *point b* (the job positions combined on the cessation date presenting a conflict on the date of activation in the sources), it is studied below how this choice may have led to a potential underestimate of the duration of the positions in the integrated base. Looking initially at the positions gathered on the closing date for which the date of CC activation was chosen (2.12%), the possible underestimate in the duration of the position compared to that indicated in LR concerns 43.3% of the positions subject to correction, but for most of them (32.7%) the date of activation present in LR is before 2016 so imputed thought tax source and it has probably been corrected to update it with the most recent CC date (any active position previously is excluded from the database). Looking now at the merged job positions on the closing date for which the date of intake of LR was chosen (0.48%), the possible underestimate in the duration of the position compared to that indicated in CC concerns 54.2% of the positions submitted for correction, and for most of them (40%) as previously seen, the date present in CC is after 2016.

1. an improvement in the estimate of the activation date for all cases of positions activated before 2014 (for which the LR uses the date present in the tax source, which is arguably less precise than the UniEmens);

2. an improvement in the identification of short-term relationships that were collapsed into a single position in the LR;

3. an improvement in the estimation of cessation dates, sometimes not declared in the sources used in the LR and not corrected in the editing and imputation process.

The analysis of the effect of the integration of the CC source on job flows shows, as expected, a large increase in the estimated number of activations and cessation. On average on the four years considered the first increases of 18.5% and the second of 20.5% (Table 4.7). The effect on the stocks of employment, is more limited, since the main issue of the LR is that two or shorter job positions between an employer and an employee are misrepresented as one longer job position. In substituting the long job position of the LR with two or shorter job positions, accounted for in the CC, the measure of the employment stock, that is the overall labour input, does not change much. It is also interesting to notice that while there is an increment on the level of the employment stock, the difference in the change of it over two consecutive years is negligible, especially when we exclude the change due to the business demographics.

Table 4.7 - Differences on employment stocks and flows between LR and extended LR - Years 2016-2019 (Percentage points) (a)

	Years				
	2016	2017	2018	2019	2016-2019
	Extended LR *vs.* LR				
Job positions ceased	20.2	20.0	21.2	22.6	20.5
Job positions activated	18.0	18.1	19.2	20.7	18.5
Stock at 1/01 year t (a)	1.2	1.6	2.3	3.0	2.3
Stock at 1/01 year t+1 (b)	0.7	1.2	1.6	2.3	0.7
Changes (a-b)	-0.5	-0.5	-0.6	-0.7	-1.7
	Extended LR *vs.* LR - only for employers always active in the period				
Job positions ceased	19.9	19.4	19.6	21.1	20.0
Job positions activated	17.2	17.9	18.0	20.0	18.7
Stock at 1/01 year t (a)	0.5	0.5	0.3	0.5	0.3
Stock at 1/01 year t+1 (b)	0.4	0.8	0.5	0.3	0.7
Changes (a-b)	-0.2	0.4	0.3	-0.2	0.5

Source: LR and Extended LR to CC

(a) Positions active on a given date include all positions activated or closed on that day.

5. The projection of Labour Register microdata based on the integration of the Compulsory Communications

As reported in Paragraph 1, since the information related to all the months of year *t*-1 of UniEmens and the other social contributions sources are available at Istat in March of year *t*, what can be done with the Labour Register, due to the timing of the production process, per se is to produce preliminary estimates of year *t*-1 with a delay of about 8-10 months. Up to August of year *t*+1 hence those are the most up-to-date estimates derived by the LR. The use of CC, whose information flows day by day into the system, may greatly improve this timeliness by projecting the set of job positions in the register referred to the year *t*-1 potentially up to a very recent time.

The methodology underlying the projection is the same described in the previous paragraphs in relation to the integration process. The projection on a CC basis can be seen in fact as the integration between the positions in the register updated to year *t*-1 and those present in the CC updated to year *t*.

The aim of this paragraph is to describe the effects of the integrated methodology on projecting the set of job positions with the use of CC up to the very last period covered by the CC source. As above mentioned, all the analyses are based on LR data 2016-2019 and CC data that cover from 2012 up to July 2020. The final scope is to project the LR referred to 2019 up to July 2020 in order to measure the most recent evolutions of employment in the first COVID-19 year. This estimation can be compared and evaluated on the basis of external data sources such as the employment estimates of the Labour Force Survey (LFS) and those of the Oros process (see also Anastasia, 2016*a*). In addition, the projections have been simulated also for the year 2019 on the basis of the LR data up to the year 2018 and CC data up to 2019. In this case, the projection can also be evaluated using all the data of the E-LR for 2019 in order to control any source of distortion due to the use of CC information not integrated with LR in year *t*. The target population of enterprises and jobs is the same one described for the integration: the set of firms belonging the sections B to S of classification NACE Rev 2, with the exceptions of temporary agencies, and all employee jobs in the target population of enterprises, with the exceptions of jobs-on-call. Operationally, the basis of the projection is the set of job positions in the E-LR belonging to the enterprises *active with employees* at year *t*-1, according to the BR referred

to that year (set A). Moreover, to provide a full account of the employment changes, the projections include the set of job positions belonging to the enterprises included in the target population that have become *active with employees* in year *t* (that is those actually born in year *t* with at least one employee or those that have passed from not having employees in year *t*-1 to having employees in year *t*) (set B). In the absence of the BR for year *t*, the *newborn* enterprises are identified thanks to information from the Oros process[14].

The application of the integration methodology in projecting employment up to year *t* (each job position is updated to the last available state in year *t* using the CC information only) implies that:

a. all job positions still active at the end of time *t*-1 in the E-LR that have no further information for *t* in the CC source are still considered active at the end of time *t* (since there has been no communication of cessation in CC);

b. all job positions still active at the end of time *t*-1 in the E-LR for which the CC signals a cessation at year *t* are ceased on that day;

c. all job positions activated in year *t* are accounted active up to their cessation date in year *t* or up to the end of the observation period if no cessation date has been communicated.

The methodology implies assuming that for all job positions belonging to enterprises ceased in year *t*-1 or year *t* the CC system has received communication of cessation. This assumption does not prove true as the database seems to lack the cessations of job positions for a certain number of enterprises that are ceased according to the information available in the external administrative source used by the Oros survey (in 2020 about 28%

---

14 The simulation exercise on 2019 and 2018 has, however, shown that to choose the set of companies to be involved in the projection cannot be considered only the presence of the company in BR in year *t*. In fact, the estimates obtained considering all the companies present in the BR target population in the reference year show an underestimated structural in the CC projection for both 2018 and 2019. The difference in the number of individuals declared in CC compared to those declared in LR in some companies in 2018 is due to the entry of non-new companies in the BR. When the company enters the field of observation the cut on BR in the two years generates new virtual relationships (all the company's relationships although active for years). If these relationships are cut off by BR from LR in year *t*-1 and have been active and stable for many years, they are not found in the CC data in year *t*, leading to an underestimate of the activated positions and an underestimate of variations in projection. To avoid this problem, the projection must include only the companies present in year *t*-1 that have not ceased and the *newborn* ones, without considering the statistical variations in economic activity.

of companies ceased according to Oros administrative information have still active job positions in CC). In addition, a distortion in the estimates can be seen if compared with external sources. This distortion is compatible with an under-declaration of the termination of relationships in companies no longer active with employees in 2020, an underestimation probably more problematic in 2020 and not evident in the comparisons carried out so far. In order to overcome this issue, the cessation dates have been corrected for all the job positions in companies that are no longer active (according to Oros process) and that seems to be still active in CC. The cessation date was set equal to the closing date of the company.

After this correction, the changes in 2020 approaches those estimated in LFS at least at the level of the total economy (while the differences at the level of economic activity might be due to classifications errors) and in Oros survey[15].

Table 5.1- Employment level at 1st semester of 2020 and changes of the first semester 2020 on the first semester of 2019 in the extended LR (a), LFS and Oros (b)

NACE	Stock at I semester of 2020 (thousands)	Change (%)	Stock at I semester of 2020 (thousands)	Change (%)	Stock at I semester of 2020 (thousands)	Change (%)
	Oros survey		LFS survey		E-LR	
Total (B-S excluding O)	12,602	-1.6	12,212	-1.3	12,564	-1.5
Industry and market services (B-N)	11,484	-1.7	11,572	-1.5	11,451	-1.6
Total (B-F)	4,472	-0.3	5,075	1.0	6,842	-0.6
Industry in the strict sense (B-E)	3,621	-0.5	4,235	0.2	3,736	-0.7
Manufacturing activity (C)	3,325	-0.6	-	-	3,408	-0.9
Supply of electricity, gas, steam and air con-ditioning (D)	83	-0.5	-	-	87	-0.6
Water supply, sewerage, waste management and rehabilitation activities (E)	195	1.2	-	-	210	1.7
Construction (F)	851	0.6	841	5.1	861	0.8
Total (G-S excluding O)	8,130	-2.3	7,136	-2.9	7,966	-2.1
Market services (G-N)	7,013	-2.5	-	-	6,853	-2.4
Retail sale repair of motor vehicles and motorcycles (G)	2,194	-0.1	2,061	-1.4	2,244	-0.8
Transport and storage (H)	1,033	-1.5	975	-3	1,029	-1.5
Accommodation and catering activities (I)	989	-10.5	918	-12.7	1,057	-9.6
Information and communication services (J)	506	-0.6	501	4.3	522	0.6

---

15 Furthermore, it can be observed that the variations calculated on the same domain for unweighted levels (active positions at 31 July, for example) are not significantly different from those for average levels.

Table 5.1 cont. - Employment level at 1st semester of 2020 and changes of the first semester 2020 on the first semester of 2019 in the extended LR (a), LFS and Oros (b)

NACE	Stock at I semester of 2020 (thousands)	Change (%)	Stock at I semester of 2020 (thousands)	Change (%)	Stock at I semester of 2020 (thousands)	Change (%)
	Oros survey		LFS survey		E-LR	
Financial and insurance activities (K)	455	-0.3	531	2.1	459	-1.4
Real estate, professional and rental activities (L-N) (c)	1,836	-2.3	-	-	1,540	-1.2
Of which: job-on-call	296	-7.8	-	-	-	-
Education, health and social work, arts and other service activities (P-S)	1,118	0	-	-	1,113	-0.5

Source: Extended LR to CC, Labour force survey (LFS), Oros process
(a) Changes calculated on the basis of E-LR refer to positions weighted by their duration in the period to make them comparable with those on the basis of Oros survey.
(b) Oros and LFS survey data are ad hoc elaborations carried out in order to make the different sources comparable.
(c) Job-on-call and job positions in temporary agencies are not included.

In order to summarise the methodology adopted, the main steps and choices are synthetically listed in the following table/prospect.

Figure 5.1- Main methodological rules of LR and CC integration to develop the Extended LR (E-LR) and its projection
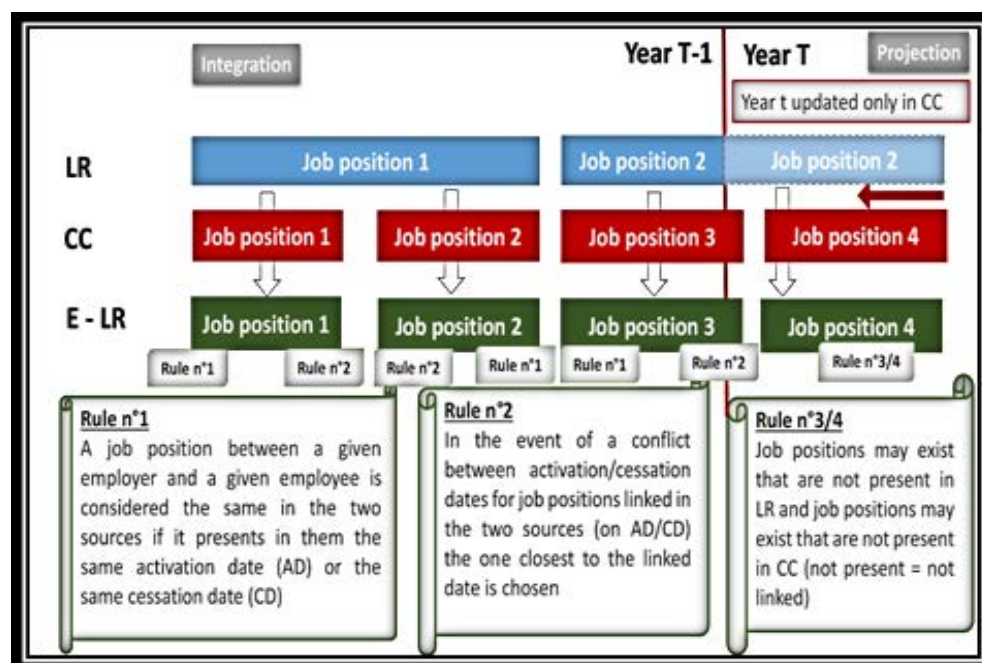
6. Measuring stocks, flows and changes in employment: the evolution of the labour market in the first part of the pandemic crisis

The procedures of integration of CC data into the Labour Register and the projection of job positions, described in the above paragraphs, enable to analyse the stocks and flows dynamics up to a very recent period measured with the precision of daily intervals.

Notice that this full stock-flows accounting of job positions is not possible with any of the existing data sources. The Labour Force survey, in fact, can only represent the worker dynamics and not the job position dynamics and not with the precision of daily intervals. Moreover, being a sample survey, these estimates would be accurate only for quite large aggregates. As for the CC data, for instance in the form of the CICO dataset, they can provide a daily account of the job flows and of the change of the job stocks, but not of the level of the job stocks.

In this paragraph, in order to provide an example of the analysis possibilities, the database built is used to study the evolution of the labour market up to July 2020, using the LR data up to December 2019 and the CC data up to July 2020.

But before describing the results it is useful to recollect the basic relationships in stocks and flows accounting (Baldi *et al.*, 2018; Anastasia, 2016*b*). In any day, $t+1$, the number of job positions (at the end of the day), is equal to the number of job positions (at the end of the day) of day t plus the activations of jobs recorded in day $t+1$ minus the cessations of jobs recorded in day t[16]. In formula:

$$J_{t+1} = J_t + A_{t+1} - C_t \tag{1}$$

The change in the number of jobs between two consecutive days is thus equal to the number of activations of day $t+1$, minus the number of cessations of day t.

$$\Delta J_{t+1} = A_{t+1} - C_t \tag{2}$$

---

16 The difference in timing between Activations and Cessations depends on the fact that while the activations are counted as active job positions from the very day in which they are recorded (that is these jobs start to be active from the activation day), the cessations are still in the count of jobs in the day they are recorded and are no longer in the set of active jobs from the day after.

By recursively substituting in [1] it is easy to obtain the relationship over a given period of time. For instance over a year:

$$\Delta J_{(t,t+365)} = J_{t+365} - J_t = A_{(t+1,t+365)} - C_{(t,t+364)} = \sum_{i=1}^{365} A_{t+i} - \sum_{i=1}^{365} C_{t+i-1} \tag{3}$$

And the relative change over a year of any given day t can be written as:

$$\frac{J_{t+365} - J_t}{J_t} = \frac{A_{(t+1,t+365)}}{J_t} - \frac{C_{(t,t+364)}}{J_t} \tag{4}$$

That is the relative (net) change of job positions is equal to the (gross) activation rate minus the (gross) cessation rate.

In Figure 6.1, where the daily dynamics of the first 7 months of 2020 is shown compared to the same period of 2019, the number of job positions (the stock measure) shows an upward trend during 2019 (also in the months not shown in the graph) and the beginning of 2020. The average stock for the first two months of 2020 is indeed higher than the equivalent figure for 2019 of about 95 thousand job positions. Starting from March 2020, when due to the policies of lock-down the economic effect of the pandemic crises has begun, there has been a sudden trend reversal that rapidly brought the daily number of job positions below the respective figure of 2019. The maximum distance is recorded for the 14th of June when the figure for 2020 is 484 thousand job positions lower than that of 2019, with a value of -3.7% in terms of per cent change. In the subsequent period there has been a slow but progressive recovery of the trend up to the end of the period, the 31st of July when the level of job positions reaches 13 million with a difference of -294 thousand (-2.2%) with respect to the same day of the year before (Baldi *et al.*, 2020).

The study of the job flows in the period January – July, which lie behind the evolution of the stocks shows that both activations and cessations started decreasing in the last days of February 2020. In particular, the drop in the levels of job positions is driven by an accelerated decrease of the job activations whose speed diminished only by mid-June. This dynamic of activation of new jobs is likely due the immediate worsening of the economic expectations of the enterprises at the outset of the crisis. The cessations also dropped but at a much lower rate due, on one side, to the policy of prohibition of dismissals that halted all termination of permanent jobs not due to the will of the workers and, on the other side, to the reduced inflows of short-term jobs explained above.

The availability of a Register that exhaustively covers the whole population of jobs in the private sector allows to break down the analysis, on various dimension, at a very detailed level. Figure 6.2 plots, at the level of 3-digit NACE sectors, the per cent change in the number of job positions against the 2019 job turnover rate (equal to the sum of the job activation rate and the job cessation rate). The figure very clearly shows that the sectors most heavily hit by the crises are those with a very high job turnover such as some activities of *Accommodation and food services*, some sectors of the *Information and communication services and Arts entertainment and recreation*. Besides having been particularly targeted by the lockdown policies, these sectors show the most dramatic decrease in the number of job positions due to their very large share of fixed-term contracts for which providing an on-the-job protection policy was not possible.

Figure 6.1 - Daily stocks of job positions and year-to-year daily changes – total, activations and cessation rate. January-July 2019 *vs.* January-July 2020
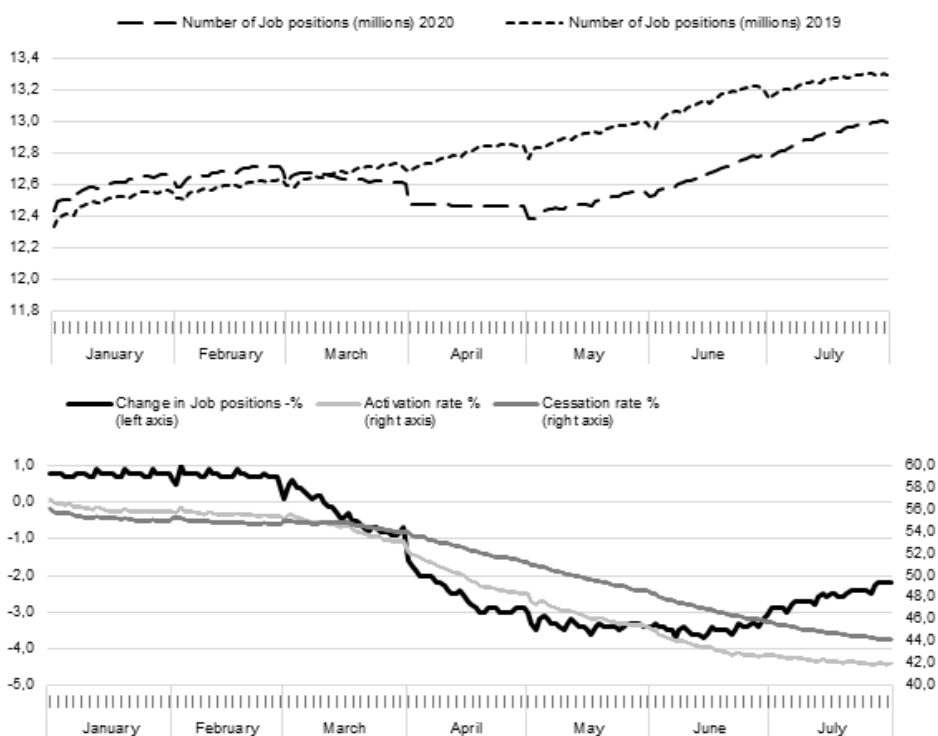
Source: Extended LR to CC

Figure 6.2 - Job positions net change rate (January-July 2020 vs. January-July 2019) vs. Gross turnover rate of 2019 (a)
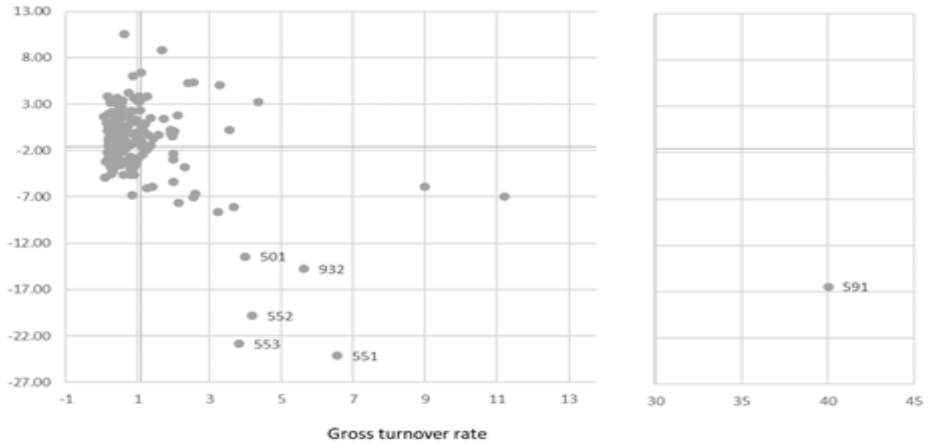
Source: Extended LR to CC
(a) The X-axis is reported with a break in it. This gap is necessary to show the exceptional outlier value of the NACE group 591 without compressing the scale of the axis.

7. Concluding remarks

The experimentation performed in this work has allowed the integration of two of the main sources for the analysis of the labour market in Italy, extending the Labour register to include Compulsory Communications.

For the very first time in Italy and, to our knowledge, in the whole world, it was built a very timely employer-employee register from which it is possible to derive a full stocks-flows accounting of employment in terms of jobs. The database has been used to shed a light on the evolution of the labour market in the first part of the COVID-19 crisis in 2020 allowing us to display the exact daily timing of the crisis not only in terms of absolute changes in the number of jobs but also in terms of relative numbers and highlighting the contribution of gross flows on net flows. Moreover, it has uncovered, at a very detailed breakdown in terms of NACE sectors, the relationship and the probable causes between the job turnover and the reduction in employment. These analyses are only examples of the potential of such a database. A number of more refined studies are possible by using the characteristics of the employer, the employee and the job and breaking them down into finer classes.

In brief, the integration has allowed reaching two main purposes: the correction of the LR for the very short duration job positions and the projection of the job stocks and flows up to the last date available in the CC source. Regarding the first purpose, in the considered period 2016-2019, a total of about 3,648,875 job positions (11.85% out of the total) were added to the LR.

As explained in the text, since the bulk of this addition is likely to consist in fragmenting in job positions of short duration what has beforehand accounted as long job positions, it does not affect much the level of average labour input measured in a year by the LR (in the period the average increase is of 1.7%). Regarding the second purpose, considering the very low number of records for which at least one edit rule failed, the projection is expected to be quite accurate. Moreover, the comparison with the dynamic shown by other stock indicators is very encouraging.

The methodology can be improved in order to try to lessen even further the micro-integration errors also with the application of the longitudinal E&I corrections of the dates of activations and cessations to the Extended LR.

Moreover, the difficult challenge constituted by the corporate transformations may be enhanc ed in order to align the two input sources before the linkage procedure.

A number of new analytical challenges can be taken up.

Firstly, it can be studied whether it is possible to project other measures of labour input, such as those traditionally calculated in the Labour Register.

A second and very interesting area of the study refers to the possibility to pass from stocks and flows accounting in terms of jobs to one in terms of workers in a unified framework.

Thirdly, with the enlargement of the LR, extending the exercise and the analysis to other sectors starting from the public one will be possible.

The fourth domain for future work is the integration of CC variables in the LR and the use of the first source to correct the information of the statistical register.

References

Anastasia, B., L. Bertazzon, M. Gambuzza, e M. Rasera. 2016*a*. "Grammatica delle Comunicazioni Obbligatorie /4. Guida ai confronti con le altre fonti statistiche sul mercato del lavoro". *Working Paper*, N. 4. Venezia, Italy: Veneto Lavoro - Osservatorio & Ricerca.

Anastasia, B., G. Emireni, M. Gambuzza, S. Maschio, e M. Rasera. 2016*b*. "Grammatica delle Comunicazioni Obbligatorie /3. Guida alle elaborazioni a partire dai dati di flusso". *Working Paper*, N. 3. Venezia, Italy: Veneto Lavoro - Osservatorio & Ricerca.

Arts, K., B.F.M. Bakker, and E. van Lith. 2000. "Matching administrative registers and household surveys". In Boon, M., P.G. Al, B.F.M. Bakker, and L. Hoeksma (*Eds*.). "Integrating administrative registers and household surveys". *Netherlands Official Statistics,* Volume 15, Summer 2000, *Special Issue*: 16–22.

Bakker, B.F.M. 2011. "Micro-integration. State of the art". In *ESSnet on Data Integration. WP1 - State of the Art on Statistical Methodologies for Data Integration*: 77-107. Luxembourg: Eurostat.

Bakker, B.F.M., and P.J.H. Daas. 2012. "Methodological Challenges of Register-based Research". *Statistica Neerlandica*, Volume 66, Issue 1: 2-7.

Baldi, C., C. Ceccarelli, S. Gigante, and S. Pacini. 2018. "The development of the Italian Labour register: principles, issues and perspectives". In Abbruzzo, A., E. Brentari, M. Chiodi, and D. Piacentino (Eds.). *Book of Short Papers SIS 2018* (49th Scientific Meeting of the Italian Statistical Society, Palermo, Italy, 20th – 22nd June 2018): 514-522. London, UK: Pearson.

Baldi, C., G. De Blasio, G. Di Bella, A. Lucarelli, and R. Rizzi. 2014. "Turning the Compulsory Communication Data into a Statistical System". In Crescenzi, F., and S. Mignani (*Eds*.). *Statistical Methods and Applications from a Historical Perspective. Selected Issues*: 217-226. Cham, Switzerland: Springer International Publishing, *Studies in Theoretical and Applied Statistics, Selected Papers of Statistical Societies*.

Baldi, C., S. Gigante, S. Pacini, e R. Rizzi. 2020. "Approfondimento: La dinamica integrata di stock e flussi occupazionali nelle imprese dei settori dell'industria e dei servizi". In Ministero del Lavoro e delle Politiche Sociali –

ML, Istituto Nazionale di Statistica – Istat, Istituto Nazionale della Previdenza Sociale – INPS, Istituto Nazionale per l'Assicurazione contro gli Infortuni sul Lavoro – INAIL, e Agenzia Nazionale per le Politiche Attive del Lavoro – ANPAL (*a cura di*). *Il mercato del lavoro 2020. Una lettura integrata*: 57-62. Roma, Italy: Istat. https://www.istat.it/it/archivio/253812.

Baldi, C., D. Ichim, F. Pintaldi, M.E. Pontecorvo, F. Rapiti, and R. Rizzi. 2018. "Digging into labour market dynamics: toward a reconciliation of stock and flows short-term indicators". In Abbruzzo, A., E. Brentari, M. Chiodi, and D. Piacentino (Eds.). *Book of Short Papers SIS 2018* (49th Scientific Meeting of the Italian Statistical Society, Palermo, Italy, 20th – 22nd June 2018): 523-530. London, UK: Pearson.

Batini, C., and M. Scannapieco. 2006. *Data Quality: Concepts, Methodologies and Techniques*. Berlin, Germany: Springer, *Data-Centric Systems and Applications*.

Daas P.J.H., and S. Ossen. 2011. "Report on methods preferred for the quality indicators of administrative data sources". *Deliverable 4.2*, Dissemination level: PU. BLUE-Enterprise and Trade Statistics - BLUE-ETS. Brussels, Belgium: European Commission, *European Research Area*.

Daas, P.J.H., S. Ossen, R. Vis-Visschers, and J. Arends-Tóth. 2009. "Checklist for the Quality evaluation of Administrative Data Sources". *Discussion paper* (09042). The Hague, The Netherlands: Statistics Netherlands – CBS.

European Statistical System Committee – ESSC. 2014. "ESS Vision 2020. Building the future of European statistics". *21st Meeting of the European Statistical System Committee*, Luxembourg, 14th and 15th May 2014.

Eurostat. 2003. "Item 4.2: Methodological Documents. Definition of quality in statistics". *Working Group* "Assessment of quality in statistics". Sixth meeting, Luxembourg, 2-3 October 2003.

Fellegi, I., and A. Sunter. 1969. "A theory for record linkage". J*ournal of the American Statistical Association*, Volume 64, N. 328: 1183–1210.

Rosén, B., and E. Elvers. 1999. "Quality Concept for Official Statistics". In Kotz, S., C.B. Read, and D.L. Banks (Eds.). *Encyclopedia of Statistical Science. Update Volume 3*: 621-629. Hoboken, NJ, U.S.: John Wiley and Sons.

Statistics Canada. 2002. *Statistics Canada's Quality Assurance Framework*. Ottawa, Ontario, Canada: Statistics Canada.

United Nations Economic Commission for Europe - UNECE. 2007. *Register-based statistics in the Nordic countries. Review of best practices with focus on population and social statistics*. Geneva, Switzerland: United Nations.

Van der Laan, P. 2000. "Integrating administrative registers and household surveys". In Boon, M., P.G. Al, B.F.M. Bakker, and L. Hoeksma (*Eds*.). "Integrating administrative registers and household surveys". *Netherlands Official Statistics*, Volume 15, Summer 2000, *Special Issue*: 7–15.

Wallgren, A., and B. Wallgren. 2014. *Register-based Statistics: Statistical Methods for Administrative Data, 2nd Edition*. Hoboken, NJ, U.S.: John Wiley and Sons, *Wiley Series in Survey Methodology*.

Wang, R.Y., and D.M. Strong. 1996. "Beyond Accuracy: What Data Quality Means to Data Consumers". *Journal of Management Information Systems*, Volume 12, N. 4: 5-33.

Zhang, L.-C. 2012. "Topics of statistical theory for register-based statistics and data integration". *Statistica Neerlandica*, Volume 66, Issue 1: 41-63.