

THE WORK PATH OF ITALIAN GRADUATES USING INTEGRATION AMONG DIFFERENT SOURCES

Claudio Ceccarelli, Stefano De Santis, Francesca Petrarca

1. Introduction

The inclusion of the Italian young graduates in the labour market is a topic that has been much discussed recently due to the various labour market reforms.

In this paper, we propose a preliminary study of this topic using data obtained from the merging among administrative archives and sample survey. A general plan of modernization which includes the Italian National Institute of Statistics (Istat, 2015c; Barcaroli et al., 2014) and other National Statistical offices has the purpose to produce the best possible estimates to meet user needs from multiple data sources, from surveys, administrative archives and new sources such as big data, and moreover to reduce burden and costs. Of course the problem of the integration of administrative archives to produce useful statistics is an issue also addressed and discussed in the rest of the world, (see Citro, 2014).

Until now, many experiments were carried out that considered different segments of the Italian population with high qualifications, e.g. high school graduates, graduates, phd., (see Petrarca, 2014; Unico Group, 2015).

The employment problem for graduates lies not only in the difficulty of getting a job but also in the quality of the demand expressed by the Italian system of goods production and services (Ciriaci et al. 2011). A widespread problem in this period is the nature of short-term job contracts very often characterised by low professional qualifications and which are not coherent with the educational curriculum, as shown in the standard sample surveys (see Almalaurea 2015, Censis, 2012; Istat, 2015a; Istat, 2015b; Capecchi, 2012). These studies gain benefits from the characteristics of administrative data and allow us to develop improvements in statistical methodologies.

2. Data sources

2.1. Education database

Ministry of Education University and Research (MIUR) delivers the archive *National Register of Students*, which provides: a National registry of university students and students of schools, databases of school structures and university structures, and of university and school teaching staff. In particular, to the aim of education studies MIUR provides, for each individual, personal data and its university career from enrolment to graduation.

In this paper, we consider all the graduates who got their university degree in the year 2011. Table 1 reports the number of graduates recorded in the year 2011 for different levels of Italian University degrees.

Table 1 – Number of graduates in the 2011 for different levels of Italian University degree

	Number of graduates
All	293,218
First cycle degree (bachelor)	167,287
Second cycle degree (master)	86,384
Single-cycle master degree	40,547

2.2. The survey-system on the study-to-work-transition

The Istat carries out the sample survey on university graduates' vocational integration by interviewing a sample of graduates who attained the university degree four years before.

Table 2 – Percentage of Italian graduates who work at the moment of the degree, after one year and after four year for different levels of Italian University degree

	At the moment of graduation	1 year from graduation	4 years from graduation
First cycle degree (bachelor)	28.7	37.4	72.8
Second cycle degree (master)	34.7	55.7	84.5
Single-cycle master degree	27.0	40.3	80.3

The survey, aiming at detecting graduates' employment conditions about four years after graduation, is part of the survey-system on the study-to-work-transition, which also includes the survey on doctorate holders' vocational integration (Istat, 2015c; Istat, 2015d) and the one on the upper secondary school graduates' transition to university and labour market (Petrarca, 2014).

A large number of information is gathered through the survey: educational experience, access to the labour market, job search, family situation, referring both to the family of origin and to the current one at the time of the interview. For more details see (Istat, 2015a) and (Istat, 2015b).

Table 2 shows that 28.7% of the first-cycle degree graduates already work at the time of the graduation and after the first year this figure is 37.4%. Very different situation for the second cycle master degree graduates who, one year after graduation, are employed in 55.7% of cases.

2.3. Compulsory Communications

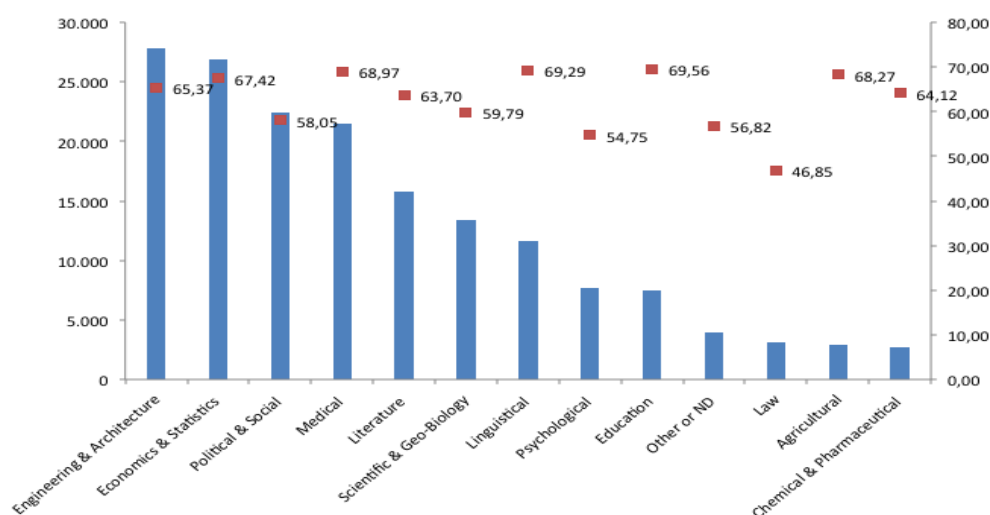
Compulsory Communications (CO) is an archive containing the information that public and private employers are required to provide to the Italian Ministry of Labour and Social Affairs at the beginning of the employment relationship or at the time of the extension, transformation and termination of the relationship of work, in agreement with the Ministerial Decree of the 30th October, 2007. The CO information system manages the information flow of administrative data so that the data can reach the Central National Coordination in real time. The information flow goes through specific peripheral systems that are delegated to data collection. The system is able to trace the flows linked to the regular employment relationships of subordinate and para-subordinate jobs. These flows are regulated within the current legislation concerning Italian citizens, and also include foreigners who have valid residence permits, even if in the country only temporarily. The system does not record, following current regulations, particular employment relationships, such as: judges, lawyers, state prosecutors, military personnel, the police force, the diplomatic corps, the mayor, and people in high administrative positions in public and private companies (e. g. presidents, managing directors). The transformation of the administrative information contained in the CO system, into statistical information is a complex task that has been entrusted to a special technical working group made up of experts from the Italian Ministry of Labour, Istat, ItaliaLavoro and ISFOL. The normalization of the archive, which is still in progress, has produced a preliminary version of the new archive called the Statistical Information System of Compulsory Communications (SISCO). It needs to be pointed out here, that all the supply job contracts communicated by Italian employment agencies and all employment relationships

involving individuals registered in the ‘Boat People’ lists have so far been excluded from this archive.

3. Data linkage

This work is a continuation of a previous paper (Ceccarelli et al., 2015), based on data coming from the *Integrated Base of Administrative Sources* (Istat), where the comparison between administrative and sample survey data is presented and moreover the problem of the definition of *worker* is illustrated.

Figure 1 – Italian first cycle degree in 2011 distribution for disciplinary sectors: the scale to the right shows the percentage of matching graduates while the one to the left shows the numbers of graduates.



In this paper, the employment condition of the graduates is obtained from the information coming from the CO.

We define *worker* a graduate for whom had at least one occurrence recorded in the CO in the observation field.

We consider the population of Italian graduates with the degree obtained in 2011 (see Table 1), in order to study their employment status in the years: 2012, 2013, 2014, 2015 and 2016.

To this aim, the dataset contain all the graduates in the 2011 is merged with the CO dataset, in the following we call this dataset *working* dataset.

The data consist of 180,592 statistical units, because among the 293,218 Italian graduates of the population potentially subject to analysis, we considered only 180,592 alumni (61.59% of the total) who had at least one occurrence recorded in the CO in the observation field (we called these graduates *matching* graduates). An occurrence is defined either by a contract which started in the period between 2012 and 2016, or by a contract which already existed at the time of graduation but which terminated in the observed period. The 112,626 alumni who do not match the CO archive are alumni who went abroad, self-employed alumni with a VAT number, unemployed alumni etc.

Figure 2 – Italian second cycle or single cycle degree in 2011 distribution for disciplinary sectors: the scale to the right shows the percentage of matching graduates while the one to the left shows the numbers of graduates.

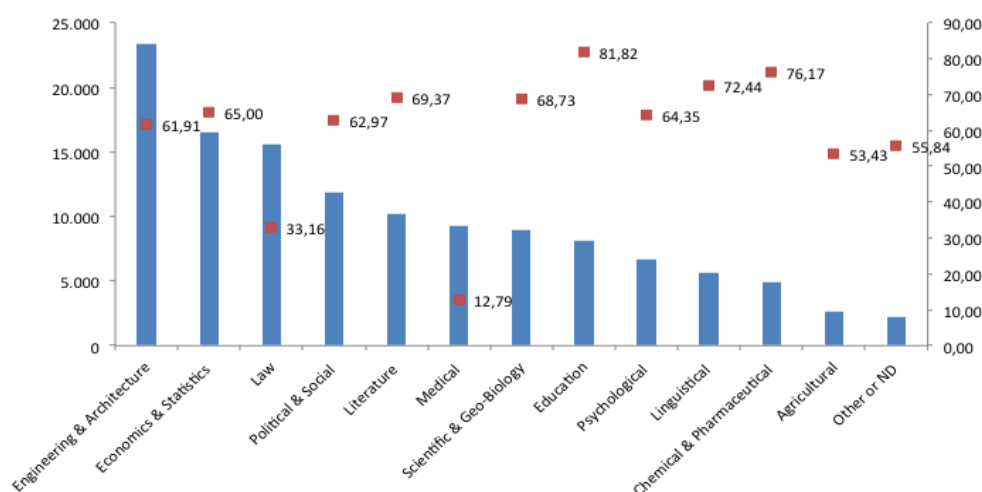


Figure 1 reports Italian first cycle degree in 2011 distribution for disciplinary sectors: the scale to the right shows the percentage of matching graduates (red squares) while the one to the left shows the numbers of graduates (blue columns). The Education, Linguistical, Medical and Agricultural are the disciplinary groups with the highest percentage of matching graduates, while the Law disciplinary group is the one with the lowest percentage of matching graduates. This is reasonable because they can be self-employed alumni with VAT number.

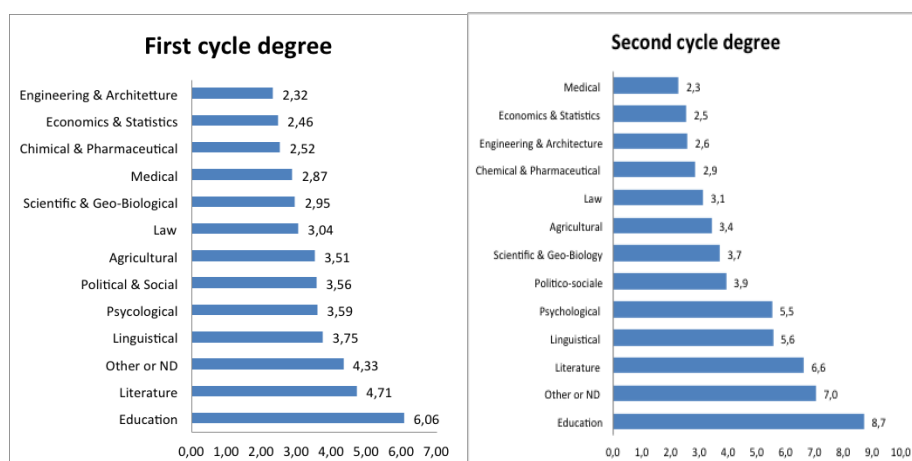
A similar representation for the distribution the Italian second cycle or single cycle degree is reported in Figure 2. The Education is the disciplinary group with the highest percentage followed by the disciplinary groups

Chimical&Pharmaceutical and Linguistical. While the disciplinary groups with the lowest percentage are Law and Medical. In Figure 3 are reported the distribution of the average number of contracts regarding the matching graduates in 2012-2016 for different levels of Italian degree.

The disciplinary groups with the highest average number of contracts are the humanities: Education, Literature, Linguistical, Psychological and Political & Social. A reasonable explanation of this fact is that these graduates have a lot of contracts of short duration because they are all graduates who work in teaching and who are called for school supplants or for contracted lectures.

On the other hand the disciplinary groups with the lowest average number of contracts are Engineering & Architettura and Economic & Statistics because these graduates have contracts that are presumably long-term or permanent positions.

Figure 3 – Average number of contracts regarding the matching graduates in 2012-2016 for different levels of Italian degree



The working dataset is merged with the statistical sample of the graduates in the 2011 (58,400 units) with the aim to study the probability of signing an optimal contract in the years following graduation. The dataset thus obtained is composed of 34,322 units.

In the following of this paper, a few logistic regression models have been investigated to study the probability of signing an optimal contract in the first years following graduation.

4. Models for the possibility of signing optimal contracts

In this Section, we report the first statistical study based on these data with the aim of investigating the integration of Italian graduates in the 2011 into the labour market as a function of their educational curriculum and of a few variables representing educational career and personal data. The results of two logistic regression models (M1 and M2) designed in order to study the probability of signing an optimal contract in the three years after graduation will be discussed. An optimal contract is a contract which has the following features: permanent position, highly qualified position, identified with the ISCO1 and ISCO2 classification, and actual duration more than or equal to 8 months (UNICO Group, 2015).

These analyses are based on a dataset reduced from 34,322 to 17,552 statistical units where we consider only the graduates belonging to: Engineering, Scientific, Literature, Economics and Statistics, Psychological, Chemical & Pharmaceutical, and Architecture disciplinary groups. This reduction was made in accordance with what was done in a previous work (UNICO Group, 2015) in order to reducing the variability of the sample.

Table 3 – Main results of the two logistic models

Model	Variables	Number of variables		Test di Hosmer Lemeshow	
		Start	Final	Value	Sign
M1	Personal Data: gender, age at the time of university enrolment, marital status, age at graduation, working condition of parents	50	6	10.37	0.24
M2	Personal data + educational curriculum: personal data+ disciplinary sector, levels of Italian graduation, degree mark, type of high school, high school rating	71	15	11.66	0.17

The differences among our models are due to the variables considered as explanatory. At the beginning of the selection procedure, M1 contained $n = 50$ explanatory variables that represent the personal data and the working condition of the parents while M2 contained $n = 71$ explanatory variables, the same as M1 plus variables concerning the educational curriculum, see Table 3.

For the M1 model the backward procedure reached an equilibrium after 4 steps obtaining a final model with 3 variables. In the case of M2 the backward selection procedure reached an equilibrium after 6 steps reducing the number of variables from 71 to 15.

Table 4 – Main results of the M1 logistic models obtained for the probability of signing an optimal contract.

Variables	Coef.	S. E.	Test W	Sign. W.	exp(coef.)
Age at the time of enrolment from 22 to 23	0.298	0.038	7.750	0.000	1.348
Age at the time of enrolment from 24 to 25	0.159	0.050	3.200	0.001	1.173
Marital status: Married	0.109	0.053	2.050	0.040	1.115
Self-employment of the mother: employed high/medium qualification	0.093	0.045	2.090	0.036	1.098
Age at the time of enrolment from 20 to 21	-0.178	0.056	-3.180	0.001	0.837
Intercept	-0.392	0.036	-10.920	0.000	0.676
Gender: female	-0.619	0.033	-18.560	0.000	0.538

Table 5 – Main results of the M2 logistic models obtained for the probability of signing an optimal contract.

Variables	Coef.	S. E.	Test W	Sign. W.	exp(coef.)
Age at the time of enrolment from 36 and over	0.611	0.187	3.260	0.001	1.843
Age at the time of enrolment from 30 to 35	0.600	0.171	3.510	0.000	1.822
Engineering disciplinary group	0.294	0.050	5.910	0.000	1.341
Marital status: Married	0.187	0.054	3.460	0.001	1.205
Self-employment of the mother: employed high/medium qualification	0.139	0.047	2.950	0.003	1.149
Age at the time of enrolment from 22 to 23	0.102	0.045	2.270	0.023	1.108
Self-employment of the mother: employed high qualification	0.087	0.044	1.980	0.048	1.091
Intercept	0.028	0.071	0.400	0.691	1.029
Not in time	-0.105	0.036	-2.930	0.003	0.901
Second cycle degree	-0.209	0.067	-3.140	0.002	0.811
gender: female	-0.363	0.036	-10.230	0.000	0.695
Economic & Statistics disciplinary group	-0.542	0.053	-10.320	0.000	0.582
First cycle degree	-0.574	0.063	-9.100	0.000	0.563
Psychological disciplinary group	-0.746	0.092	-8.090	0.000	0.474
Architecture disciplinary group	-0.944	0.077	-12.290	0.000	0.389
Literature disciplinary group	-1.313	0.072	-18.270	0.000	0.269

In Table 4 and Table 5 a summary of the main results of each of the two final models has been reported. These Tables contain the names of the significative variables, the coefficient (parameter β of the logistic regression) that represents the expected amount of change in the logit for each one unit change of the predictor, the standard error of β (S.E), the value of the Test Wald, its p-value (Sign.W.) and exp(coeff.).

For each resulting model, all variables have the parameters β significantly different from zero (Sig. $W < 0.05$) and the variables are listed in descending order of $\exp(\beta)$ in order to allow the verification of the importance of each variable that characterizes the presence of an optimal contract. In the case of M1, the variable "Age at the time of enrolment 22-23" and "Age at the time of enrolment 24-25" have the highest values of the coefficients ($\beta = 0.30$ and $\beta = 0.16$ respectively). Therefore, for example, age at the time of enrolment 22-23, the odds of signing an optimal contract changes by a factor of 1.35, holding all the other variables at fixed values. For the M2 model, the variable with the highest impact is the one representing the alumni who are no longer very young at the time of enrolment at university.

Comparing the results obtained by the two models, it is worth noting that the variables connected to the disciplinary sectors Engineering, Economic & Statistics, Psychological, Architetture and Literature, which are not included in the M1 model, appear to be decisive in order to get an optimal contract. The goodness of fit of all our models to the data, as assessed with the Hosmer-Lemeshow test, shows that the models in the first and last steps of the procedure backward, have a good fit to the data since the significance values are all greater than 0.05.

5. Remarks

We underline the importance of using administrative data for the study of the entrance of graduates into the Italian labour market. We have briefly shown that an administrative archive is flexible and rich enough to analyse the work paths of graduates in the years following graduation.

One of the results of the data matching is that the disciplinary groups with the highest average number of contracts are the humanistic because these graduates have a lot of contracts of short duration.

On the other hand, the disciplinary groups Engineering & Architetture and Economic & Statistics have the lowest average number of contracts that are presumably related to long-term or permanent positions.

Moreover, the administrative data allow us to study the evolution of graduates after the graduation and therefore to analyse changes in their job position.

We presented the preliminary results of binary models based on syntheses indicators.

The choice of the disciplinary sector is decisive for the achievement of the optimal contract.

The analyses on administrative database could be adopted as a permanent monitor of the entrance of graduates into the Italian labour market over the years which may be accompanied from the sample survey in order to enrich the dataset with information which can not be obtained from administrative data. In this way, it is possible to reduce the economic burden of the sample survey.

References

- ALMALAUREA 2015. *XVII rapporto 2015 AlmaLaurea sulla condizione occupazionale dei laureati*. Tech. rep., Consorzio Universitario AlmaLaurea
- BARCAROLI G., FALORSI P.D., FASANO A., MIGNOLLI N. 2014. A Business Architecture Model to Foster Standardisation in Official Statistics. *European Conference on Quality in Official Statistics - Q2014*
- CAPECCHI S., IANNARIO M., PICCOLO D. 2012. *Modelling job satisfaction in AlmaLaurea surveys*. AlmaLaurea Working paper n 50
- CARPITA M 2011. *Laureati Stella: Rapporto statistico 2008-2011- Progetto CILEA*. Grafiche Porpora
- CECCARELLI C., MONTAGNA S., PETRARCA F. 2017. Longitudinal analysis of Italian graduates employment paths through integration among different administrative sources. *Rivista Italiana di Economia, Demografia e Statistica*, Vol LXXI, No.3 Luglio-Settembre, pp. 29-40.
- CENSIS 2012. *Quarantaseiesimo Rapporto sulla situazione sociale del Paese 2012*. Milano: Franco Angeli
- CIRIACI D., MUSCIO A. 2011. *University choice, research quality and graduates' employability: Evidence from Italian national survey data*. AlmaLaurea Working paper n 48.
- CITRO C.F. 2014. From multiple modes for surveys to multiple data sources for estimates. *Survey Methodology*, Vol. 40 No. 2, pp. 137-161.
- ISTAT 2015a. University graduates vocational integration. <http://www.istat.it/en/archive/82425>
- ISTAT 2015b. I percorsi studio e lavoro dei diplomati e dei laureati. <https://www.istat.it/it/files/2016/09/I-percorsi-di-studio-e-lavoro-dei-diplomati-e-laureati.pdf?title=Percorsi+lavorativi+di+diplomati+e+laureati+-+29%2Fset%2F2016+-+I+percorsi+di+studio+e+lavoro+dei+diplomati+e+laureati.pdf>
- ISTAT 2015c. Indagine sull'inserimento professionale dei dottori di ricerca: informazioni sulla rilevazione. <https://www.istat.it/it/archivio/8555>
- ISTAT 2015d. Transition of doctorate holders from education to work. http://www.istat.it/en/files/2015/01/EN_doctorate-holders-

- 2014_finale.pdf?title=Doctorate+holders+from+education+to+work++-+21+Jan+2015+-+Full+text.pdf
- PETRARCA F. 2014. *Assessing Sapienza university alumni job careers: Enhanced partial least squares latent variable path models for the analysis of the UNI.CO administrative archive*. PhD thesis, University of Roma Tre, [On- line; accessed 3-Jun-2014]
- RUNCI M.C., DI BELLA G., GALIÉ L. 2016. *Il sistema di integrazione dei dati amministrativi in Istat*, Istat working paper n. 18/2016.
- UNICO Group 2015. *La Domanda di Lavoro per i laureati. I risultati dell'integrazione tra gli archivi amministrativi dell'Università Sapienza di Roma e del Ministero del Lavoro e delle Politiche Sociali*. Edizioni Nuova Cultura

Summary

The work path of Italian graduates using integration among different sources

In questo lavoro si analizzano dati provenienti dall'integrazione di differenti archivi amministrativi con l'indagine campionaria sull'inserimento professionale dei laureati al fine di studiare i percorsi occupazionali a uno, due e tre anni di distanza dalla laurea (Istat). Da questa integrazione è emersa una situazione molto complessa che deve essere analizzata e correttamente interpretata. Si presentano due modelli logistici per studiare la probabilità di ottenere un contratto ottimale nei primi anni dalla laurea. La scelta del settore disciplinare risulta essere determinante per il raggiungimento del contratto ottimale.

This paper analyses data coming from the integration of various administrative archives together with the sample survey on *university graduates' vocational integration* (Istat). From this integration a very complex situation has emerged that needs to be analysed and correctly interpreted. Two logistic regression models have been investigated to study the probability of signing an optimal contract in the first years following graduation. The choice of the disciplinary sector is decisive for the achievement of the optimal contract.

Claudio CECCARELLI, Italian National Institute of Statistics, clceccar@istat.it
Stefano DE SANTIS, Italian National Institute of Statistics, sdesantis@istat.it
Francesca PETRARCA, Italian National Institute of Statistics,
francesca.petrarca@uniroma1.it