

SESSIONE III

INTEGRAZIONE E USO DI DATI AMMINISTRATIVI PER FINI STATISTICI

L'eterogeneità dell'occupazione non regolare in Italia: alcune evidenze dai microdati del campione forze di lavoro integrato con le fonti amministrative

Carlo De Gregorio e Annelisa Giordano

Heterogeneity of irregular employment in Italy: some evidence from Labour force survey integrated with administrative data

Carlo De Gregorio, Annelisa Giordano

Istat

cadegreg@istat.it, giordano@istat.it

Abstract

The intrinsic heterogeneity of irregular employment is analysed by exploiting the microdata derived from the statistical integration of Labour force survey sample with the administrative records that trace regular jobs, where irregularity is flagged by comparing independent sources. Following previous approaches we use logistic regression to model the probability of being irregular worker as a function of individual characteristics and local context indicator. A segmentation of irregular employment shows how heterogeneous is the combination of labour supply conditions with actual labour demand. Our results give the possibility to appreciate the coexistence of different irregularity profiles obtained by combining sector and socio-economic conditions. They seem to support the adoption of an approach to active policies where local conditions should receive greater attention.

Keywords: Labour market, Irregular employment, Non-observed economy, Logistic regression, Multiple correspondence multiple, Cluster analysis

Sommario

La natura profondamente eterogenea dell'occupazione non regolare viene qui analizzata attraverso l'uso del campione dell'indagine sulle forze lavoro i cui microdati sono stati integrati con le informazioni contenuti negli archivi amministrativi che tracciano l'occupazione regolare. Attraverso un modello logistico è stata modellata la probabilità di avere un'occupazione irregolare in funzione delle caratteristiche socio-demografiche dell'individuo, di fattori locali di contesto relativi al mercato del lavoro e alla struttura produttiva, e delle caratteristiche della sua posizione lavorativa. Una segmentazione dell'occupazione non regolare evidenzia la combinazione fra caratteristiche dell'offerta e struttura della domanda ed evidenzia la coesistenza di profili eterogenei.

Parole chiave: Mercato del lavoro, Occupazione non regolare, Economia sommersa, Modello logistico, Analisi delle corrispondenze multiple, Analisi dei gruppi

Introduction

Research has recently moved the focus towards the multi-facet nature of irregular employment, contributing to enrich analysis and – potentially – to address policy¹. Heterogeneity can be expected to derive from the interaction of individual characteristics (such as education, skill, household structure, economic conditions, etc.) with the working of local economies. Additional ele-

¹ See Cappariello *et al.* (2009) and the literature review on these issues. See also the Italian version of this work: Cappariello-Zizza. 2009. *Istruzione ed economia sommersa*. In: Banca d'Italia, Mezzogiorno e politiche regionali. Seminari e convegni n. 2, novembre, p. 191-214. Following OECD manual, irregular or underground employment is meant to be "Employment concealed by the enterprises choosing not to respect employment regulations or immigration laws by hiring labour off the books". See OECD (2002, p.38).

ments of heterogeneity are induced by the fading borders between regular and irregular jobs².

Microdata are needed to study irregularity profiles. In the past, important achievements have been obtained in Italy using indirect tools, mainly founded on the cross comparison of detailed domain aggregations of employment data from independent sources³: this approach provided accurate estimates of irregular labour input according to some breakdown. More recently, the analysis of household survey microdata has gained ground, based on the indirect detection of irregularity from the survey questionnaire: these approaches, if on one side do not provide unbiased level estimates, on the other side open the way to the study of individual characteristics and segmentation⁴.

This paper develops a similar approach by exploiting the microdata from the statistical integration of Labour force survey with the administrative records (hereafter ADMIN) that trace regular jobs. This integrated sample (named LFS-ADMIN⁵) has the advantage of allowing the use of a huge amount of microdata where the detection of irregularity is derived within the integration process by a correction of level biases⁶. In what follows we give a short presentation of LFS-ADMIN with a description of irregular work estimates (par. 1); we discuss some results obtained from modelling the probability of being in irregular employment (par. 2); then we provide a segmentation of irregular employment based on the characteristics of each individual, on his job and on his traces in ADMIN (par. 3). Some conclusions are finally drawn.

1. The LFS-ADMIN integrated sample and the identification of irregular workers

LFS is a continuous survey with a yearly sample of about 600 thousands interviews representative of resident population⁷. LFS-ADMIN integration has brought to the estimate of the individual actual employment status based on a statistical reconciliation of the individual information gathered from both sources and referred to the reference week of LFS interview: this implied the adoption of methods to detect, measure and correct the biases affecting both sources⁸. Irregular jobs have been

² An important share of irregular labour input actually derives from formally regular employment: thus it is increasingly embarrassing to represent irregularity as a binary variable. Boeri *et al.* (2002) focused instead on the fading borders between irregularity and unemployment. It is increasingly evident, anyway, that the fraudulent misuse of labour contracts is increasingly evident and generates unreported working time with the partial evasion of social security and tax duties. The emersion of this “grey” economy might also be envisaged as the result of a partial adaptation to the policy actions tailored to contrast purely hidden jobs. On the relevance of grey labour input in Italy see the final report of the so called “Giovannini Commission” (MEF, 2011), and Isfol (2007b). See the huge work (mainly through empirical analysis) provided by Williams (e.g. Williams (2010)). See also, more recently, De Gregorio-Giordano (2014) on the diffusion of false part-time contracts in Italy.

³ See for example Calzaroni (2000) whose approach has been founded on the comparison between Census data and Labour force survey data. Boeri *et al.* (2002) support the idea that a large share of irregular employment is hidden among those who are classified unemployed or inactive. See Zizza (2002) for a survey. See also Cappariello *et al.* (2009) or, on a dedicated perspective, Baccini *et al.* (2003), Isfol (2007a, 2011).

⁴ Cappariello *et al.* (2009), in particular, derive very interesting results by flagging individuals in employment as irregulars if they do not declare social security coverage. They work on the microdata of the biennial Survey on Household Income and Wealth, run by the Bank of Italy with a sample of nearly 8.000 households. In spite of the limited sample size and a narrow definition of irregularity, they provide several interesting insights, inter alia on its ties with education and gender. Boeri *et al.* (2002) worked on a survey sample in Sicily where irregularity was directly asked in the questionnaire.

⁵ The methodology adopted to build LFS-ADMIN has been developed by an Istat working group and it is described in AA.VV. (2014). ADMIN data derive mainly from social security sources on employees (INPS-EMENS, INPDAP, ENPALS, INPS-DMAG, Lavoratori domestici) and self-employed (INPS-Gestione separata, INPS-Collaboratori professionali, Sistema informativo ASIA-Indipendenti, Autonomi agricoli). They have all been used to build the employment register (DB Occupazione) supporting the system of Business registers (ASIA).

⁶ The integrated sample LFS-ADMIN has been developed by Istat with reference to the two-year 2010 and 2011 with the purpose of supporting national accounts benchmark estimates of regular and irregular labour input (namely number of persons in employment, jobs and hours actually worked). See Istat. *I nuovi conti nazionali in SEC 2010*. Nota informativa, 6 October 2014 (pages 21-25) or also Istat. *Il ricalcolo del Pil per l'anno 2011*. Nota informativa, 9 September 2014 (pages 9-11).

⁷ Permanent members of collective facilities (hospices, religious institutions, barracks, jails, etc.) are excluded from LFS. Non-residents comprise foreign citizens irregularly present in Italy, who are consequently not included in this analysis: notice that the rate of irregularity in this segment is very high. National accounts estimates on the contrary are exhaustive.

⁸ The methods adopted for data integration are fully described in AA.VV. (2014). For a short outline see De Gregorio, Filipponi *et al.* (2014). Previous research by the ESSnet on data integration has been a precious guidance to the approach. See also García Martínez (2011), Hochfellner (2011), Kuijvenhoven *et al.* (2011), Linder *et al.* (2012), Zhang (2012). It is well known that LFS adheres to ILO standards that principle cover any kind of labour input, regular or irregular. On the contrary, ADMIN status is mostly referred to administrative rules and only refers to labour input with official traces; furthermore it may include false positives. Boeri *et al.* (2002), for instance, affirm that a meaningful share of unemployed and inactive according to LFS are actually employed in the informal sector. The accuracy of ADMIN is source dependent and derives on the available information concerning the time-span of actual labour input. A lack of precision affects mainly the sources on self-employment. See AA.VV. (2014) for more details.

defined as employment spells not matched with validated ADMIN signals in the reference week⁹. Integrated job data are thus combined with the personal characteristics of the worker and with the whole profile of his yearly ADMIN records. Since our focus is on employment, we have selected from LFS-ADMIN only the primary job of each individual¹⁰: considering both years together, the total sample consists of about 480.000 persons in employment, 48.000 of which are irregular¹¹. LFS weights are used for the grossing up.

LFS-ADMIN estimates for the whole period 2010-2011 confirm some expected characteristics of irregular employment highlighted by other independent estimates¹². The incidence of underground employment is estimated nearly 10% of total employment in the target population. Higher rates can be found among women, foreign citizens (especially from EU), self-employed, young people, low education segments, South, and in agriculture, constructions, hotels and restaurants, households services. Other aspects stand out clearly: elderly people and low skilled professionals seem affected by higher rates too; households structure and income play a non-secondary role.

Integration delivers further challenging hints. For instance, the contiguity between regular and irregular jobs¹³. This seems to indicate a constant switch from regularity suggesting the treatment of irregularity as a sort of continuous or at least multi-modal variable¹⁴. Although the well-known stereotypes are evidently confirmed, the integrated sample confirms also that heterogeneity lays behind them¹⁵. Irregularity is spread across many segments of the labour market usually intended as at-low-risk and this mere fact deserves some focus in order to target analysis and policies.

2. Modelling the probability of working underground

The probability of a person in employment to work underground can be analyzed from different perspectives. The individual characteristics (such as age, gender, household, citizenship, etc.) have been primarily considered as independent variables: other variables have been progressively introduced to summarize local context effects.

A battery of Nuts3¹⁶ labour market indicators is used to monitor their effects on individual risk profiles: we opted for the rates of activity and unemployment, the contact rate of job centers (be it public or private)¹⁷, the coexistence rates of the so called *grey area* with official labour force and of potential employment¹⁸ with total employment¹⁹. A partition of the Nuts3 into eight clusters - de-

⁹ The integrated dataset lists the jobs performed by individuals actually employed, with indication of the order of the job (primary, secondary, etc.), regularity status, Nace, actually worked hours, type of employment, Isco, business register data on the employer and the rest of LFS information. In the case of irregular jobs, the information is derived mostly from the answers to LFS questionnaire and from their recent regular working history recorded in ADMIN. ADMIN data can be organized longitudinally and individual regular histories can be used more efficiently to outline and detect irregularity.

¹⁰ According to the ESA, the primary job determines the characteristic of each employed, namely whether he is an employee or a self-employed, the sector in which he works and also the regular or irregular nature of the worker. This independently from the characteristics of any eventual secondary jobs.

¹¹ We have thus excluded about 55 thousands secondary jobs, 8.000 of which correspond to irregular jobs. All these figures are very similar in the two years.

¹² See for example Istat national accounts estimates (*La misura dell'occupazione non regolare nelle stime di contabilità nazionale*, <http://www.istat.it/it/archivio/39522>, or Istat. *L'economia sommersa e il lavoro non regolare*. Audizione del Presidente dell'Istituto nazionale di statistica presso le Commissioni riunite V Commissione "Programmazione economica, bilancio" del Senato e V Commissione "Bilancio" della Camera, 21 July 2005). See also Cappariello *et al.* (2009).

¹³ Underground workers are frequently traced in ADMIN in the reference year although such traces are not compatible with coverage in the reference week: they anyway characterize the working activity in other parts of the year.

¹⁴ It is also worth mentioning that an important share of irregulars declare to LFS to work in large local units. This aspects opens the way for analysis dedicated to underground subfurnitures of services to larger units. Furthermore, irregular jobs appear associated with lower actually worked hours (about 14% less than regular ones): nevertheless such difference is tiny in segments marked by a higher incidence of irregular jobs: another promising subject for further research.

¹⁵ Cappariello *et al.* (2009) also stress this point.

¹⁶ Corresponding to the more than one hundred "province" in which Italy is split.

¹⁷ This rate is computed as the share of unemployed and *grey area* inactive population (willing to work but who don't search actively or who are not immediately available for starting a new job) that contact job centers in the weeks before LFS interview. The idea behind this choice is a higher use of official channels is an indicator of active policy concern and marks an antibody against informal jobs.

¹⁸ As defined by the sum of unemployed and *grey area*.

rived from the DBGEO database developed by the tax authority²⁰ - has been used as a proxy of the local attitudes towards tax compliance. The effects of the employment by gender structure of local regular business have also been taken into account.

A last set of input variables concerning the actual job of the individual - Nace and type (employee or self-employed) - is used to focus on the demand side of the irregular labour market.

By using the nature of employment - whether regular or not - as the response variables, a logistic model has been run to model the probability of this event in function of the above sets of variables. Several specifications have been tried, changing the sets of variables, the interactions and the model groups²¹. What follows is the general simple effect version:

$$\text{logit}(IRR_{i(gk)}) = \alpha + P'_{i(gk)}\beta + X'_{i(gk)}\gamma + W'_{i(gk)}\lambda$$

where i , g and k stand for the individual i , resident in the k -th Nuts3 and whose gender is g ; IRR is the binary response variable; P summarizes social and demographic characteristics of each individual; X are the local indicators on labour market, tax compliance and business structure; finally W labels the variables describing the actual job of each individual.

We report here some results obtained with three simple effects models run separately on the two gender groups: model A uses only P variables, model B introduces the X set and model C uses also W variables. All three models appear to fit the data well, with increasing scores as we pass from model A to model C²². Both model groups show that foreign citizens have a higher probability of being in irregular employment, although EU citizens have a far larger risk of being irregular with an odds ratio in model A larger than 1.4 points compared to the rest of foreigners. This difference somehow reduces as context and job effects are introduced: in model C that ratio drastically decreases - but only for males - remarking the importance of demand side effects. Age appears characterized by some symmetry: the probability of being irregular grows as we depart from central age classes. For young people age and household effects add up, given the higher odds associated to individuals living with parents. Some gender gap emerges if we consider the effect of household structure. Men living alone have a relatively higher probability of being underground, with an odds that doubles that of adults living with a partner and a son (the benchmark less at risk). The corresponding odds ratio for women is far lower and this may be due to the conditions laying behind the choice of living alone. Another class with a higher irregularity risk is the class of single parents living with sons; here the odds nearly double the benchmark. The presence of another income earner in the household works quite differently according to whether he is regular (slightly lower risk) or irregular (much higher risk). A low education attainment is a crucially risky condition, even more for women²³.

¹⁹ These indicators have been derived for total population aged 15-64 years and for younger population (15-34 years), separately by gender. The contact rate has been derived only for population 15-64 by gender. In order to avoid the drawbacks of the strong correlation among these indicators we also used their first three principal components, estimated by gender. The principal components were extracted, separately by gender, from a dataset of 110 Nuts3 indicators without weighting. The first one (85% of total inertia) expresses the general quality of the local labour markets: high activity rates and relatively strong active policies as opposed to unemployment and grey area. The second one (8%) gathers the effects of official placement facilities in moving potential labour force from inactivity to unemployment. The third factor (3%) describes the intensity of official placement non accompanied by evident effects.

²⁰ The clustering is based on variables concerning tax behavior, criminality, consumption patterns, business structure, technological development, transport infrastructures, type of taxpayer. See "Indagine conoscitiva sugli organismi della fiscalità e sul rapporto tra contribuenti e fisco", Audizione del Direttore dell'Agenzia delle entrate, Senato della Repubblica, VI Commissione finanze e tesoro, Rome 2 April 2014.

²¹ The main results obtained through to alternative specifications do not differ substantially. Models with weighted and unweighted observations have been tested, without appreciable differences. All the data reported in this work derive from the use of weighted observations.

²² For instance, for both model groups the concordance ratio ranges from about 67-68% to 73-74%

²³ It is interesting to notice that, in the case of men, the possession of a university degree puts the individual more at risk as compared to an intermediate education (the completion of secondary schools). This aspect, however, needs further analysis in order to explain why the same is not

Labour market conditions suggest gender differences. The risk of irregularity for men grows more rapidly as labour market weakens. But as for women, active policies if associated to high unemployment rates may partially translate into a higher participation in the wrong side of the market: this does not seem to be the case for men. DBGEO clustering copes well with explaining underground work for both genders, in particular when included in the low tax compliance cluster. The effect associated with the structure of regular business deserves some attention. For men, a higher share of regular employment in “difficult” sectors (such as agriculture and construction) brings to a lower overall probability of underground jobs²⁴. On the contrary, a higher weight of microenterprises seems related to higher irregularity risks.

The introduction into the analysis of details on irregular jobs, brings into light other gender differences. While in general self-employed are more at-risk-of-irregularity, such effect is much stronger for women. Let alone jobs in agriculture, whose odds are more than twice those of industry, higher risks are run by women in household services and by men in construction and trade. Finally, while industry is the less risky sector for men, this is not so for women.

If we take as a benchmark one of the most virtuous profiles (profile 1: a middle-aged highly educated individual in a household with a regularly employed spouse and two sons. See Table 1), model A predicts for a male a 4.2% probability of being underground. Should his Isced be the lowest one, his probability would increase by 1.3 p.p., and by further 1.7 p.p. if his wife would not work. In the same situation a woman is predicted to start 2.2 p.p. higher in profile 1, and her probability would jump up more rapidly if she had low education and no income from her husband. The same individuals, living single and with a low Isced, would both show about 12% probability of being irregular. If they were EU citizens, the predicted probability would more than double.

Table 1. Predicted probability of being a person in irregular employment, by citizenship, profile and model (%)

Profile	Age	Role and Hh structure	ISCED	Other incomes	MALE						FEMALE					
					A	B		C		A	B		C			
						min	max	min	max		min	max	min	max		
ITALIAN																
1	35-54	Spouse (2 Parents & ≥2 sons)	5	REG	4.2	2.0	11.3	1.6	19.1	6.4	2.8	14.9	1.8	29.5		
2	15-24	Son (2 Parents & ≥2 sons)	2	REG	19.2	9.6	39.8	7.5	50.6	29.8	15.7	51.8	11.8	70.0		
4	35-54	Parent (1 Parent & ≥2 sons)	2	none	9.6	4.2	21.3	3.6	28.0	13.0	5.3	25.7	4.1	45.2		
5	15-24	Son (1 Parent & ≥2 sons)	2	none	25.7	11.9	44.1	9.2	54.6	32.4	17.8	49.6	13.9	65.9		
EU																
1	35-54	Spouse (2 Parents & ≥2 sons)	5	REG	10.5	6.0	25.3	4.4	33.1	15.1	7.3	31.4	5.8	37.3		
2	15-24	Son (2 Parents & ≥2 sons)	2	REG	38.7	25.5	49.6	18.8	52.1	52.3	39.6	70.8	34.8	76.3		
3	25-34	Single	2	none	31.4	19.7	58.7	15.1	65.5	33.2	21.1	55.9	15.2	70.5		
Extra EU																
1	35-54	Spouse (2 Parents & ≥2 sons)	5	REG	7.3	4.8	21.7	3.7	27.8	11.2	5.9	25.9	3.9	27.2		
2	15-24	Son (2 Parents & ≥2 sons)	2	REG	29.7	20.6	59.1	17.1	65.1	43.6	27.6	67.2	22.4	67.2		
3	25-34	Single	2	none	23.5	15.2	52.8	11.9	65.5	26.0	17.0	50.1	12.5	68.6		

Note: (a) Weight by gender and citizenship

Source: LFS-ADMIN, Years 2010-2011

Take a young individual living with both parents in a household with at least one regular income earner (profile 2): his predicted probability goes near to 20%, and nearly 30% if she was a girl. Both probabilities pass respectively 30% and 40% if the income earner was irregular. A higher Isced would reduce both probabilities and the gender distance. But if profile 2 was an EU citizen

found for women: a possible answer can be drawn from the fact that the introduction of context factors reduces this unexpected difference, and this could be interpreted as sign of the weakness of local markets and policies to meet this segment of labour supply.

²⁴ This might be connected to the emersion of previously underground activities. This effect is anyway not at work for women.

the predictions would double. A foreigner living single with a low Isced and an age between 25 and 34 years (profile 3) has between 25% and 33% probability of being irregularly employed.

A middle-aged parent living alone with at least two “not income earner” sons (profile 4) has almost 10% probability of being irregular, 13% if woman, 28% if woman and EU citizen and more than 20% if Extra EU. A slightly higher Isced level would cut the prediction.

The adoption of model B and C introduces variability in predictions. For male in profile 1 it ranges from 2% to 11% if context factors are introduced and hits a maximum at 19% with model C; for women the right tail of the distribution is longer. In general, the distribution of prediction is strongly skewed for the more virtuous profiles. The predicted probability of profile 2 for women ranges from 16% to 52% if we account for context factors, and passes 70% if demand side is considered: the same profile for men has a maximum ten point lower. Local factors generate heterogeneity also within segments apparently protected against the risk of irregularity.

3. A segmentation of irregular employment

The sub-sample of the individuals in LFS-ADMIN with an irregular primary job (nearly 25 thousands) has been analysed with a sequential use of correspondence analysis (ACM) and Ward clustering: the focus now is more strictly on the irregular job itself and on the sector specific features related to the use of underground labour input. The variables used in the analysis are those included in the sets P and W mentioned above: further variables from individual annual ADMIN traces have been added to W , scaled according to the intensity of the signals²⁵. With about 70 independent modalities, the first ten eigenvalues account for about 40% of total inertia²⁶: the first factor (6.5% of total inertia) opposes two poles that we could summarize as “*unskilled blue collars*” vs. “*skilled self-employed*”²⁷; the second factor (5.8%) draws specularly “*skilled white collars*” vs. “*low education self-employees*”²⁸.

A nine cluster partition based on the first ten factors is reported in Table 2. Construction and household services show a specialisation in the larger cluster 3, characterised by low skill part-time employee jobs: it sounds reasonable to find in this cluster a relatively strong presence of EU citizens. Construction is also represented in cluster 8, where more skilled craft professions are included and where also industry and trade have a meaningful presence: in this segment young people and foreign residents central and northern regions are relatively more frequent. Industry itself is a major presence in cluster 9 where employees have intermediate skills with traces in the annual ADMIN: northern regions and EU citizens have some ties with this group.

Agriculture has two main specialisations. Employees from this sector feed the cluster of older, low education and low-skilled workforce (cluster 6): South and foreigners describe well the segment. A very low education score although accompanied by high skill professional levels, draw cluster 4 where agriculture self-employment has a stronghold: there are relatively old, mostly Italians and from the South. Self-employment in trade is also well represented in this cluster together with cluster 1 (like business services), where education is higher and where central and northern regions, and male employment, have a relatively higher presence.

²⁵ The net monthly income declared to LFS, the hours actually and normally worked, the number of secondary jobs have been used as illustrative variables.

²⁶ Better results were obtained by running separate analyses for the main economic sectors.

²⁷ On one side, foreigners, young men, employees, low education, elementary profession, full-time; on the other, self-employed, professionals and entrepreneurs, central age classes, higher education, part-timers, also women, with extremely weak ADMIN traces.

²⁸ On one side, young women, with medium-high education, clerical workers, northern and central regions, with ADMIN traces; on the other, self-employed skilled workers, with low education, aged, men, Italians.

Table 2. Clusters of individuals in irregular employment by segment (distribution; specialization rates by segment; %)

Cluster	Cluster	Distrib. %	Specialisation rate (a)											
			Nace							Women	Young 15-34	EU	Extra EU	South
			Totale	Agriculture	Industry	Construc- tion	Trade & Horeca	Business services	Household services					
1	1. Self employees, Very weak ADMIN, Italians	7.8	100	29.4	98.0	109.8	164.5	150.2	46.9	77.3	58.8	16.9	26.5	84.2
2	2. High education, Large units, White collars, High skill, Weak ADMIN traces	9.6	100	12.8	35.8	15.5	25.8	128.0	213.5	115.4	66.8	31.4	16.7	91.0
3	3. Employees, No ADMIN traces, Part-time, Low skill, Low education	28.8	100	65.1	85.8	117.8	98.6	81.0	121.7	105.2	103.3	141.7	100.4	106.0
4	4. Self-employed, ADMIN traces, Aged, Italians, Low education, High skill	9.6	100	268.0	49.5	98.4	157.7	98.1	36.1	87.8	61.4	14.6	46.5	127.9
5	5. Young, unmarried, high education, medium-high skill, Italians, weak ADMIN	5.4	100	9.7	71.1	30.4	67.7	188.0	133.3	117.6	169.5	21.6	24.3	69.2
6	6. Employees, Elementary occupations, Low education, South, Parents, Weak traces	6.6	100	757.4	60.1	39.1	57.0	30.8	14.9	80.2	86.9	135.0	135.2	167.3
7	7. Women, Foreigners, weak ADMIN, Single, part-time	3.7	100	9.2	8.2	11.8	18.3	15.7	294.8	181.4	70.6	353.8	484.3	56.3
8	8. Weak ADMIN, Blue collars, Craft worker, Men, Medium-young, Low education	11.0	100	21.1	146.5	209.2	136.2	99.3	43.9	73.0	148.4	138.7	138.2	92.1
9	9. Blue collars, traces in ADMIN, Intermediate skill	17.6	100	46.2	199.6	108.1	102.2	100.6	69.0	99.1	105.5	94.0	120.1	92.1
	Total	100	100	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Source: LFS-ADMIN 2010-2011

Note: (a) Obtained as an index of actual frequency as compared on theoretic frequency (n/n*)

Cluster 5 shows a noticeable presence of persons employed in business and household services: people in this cluster is somehow between employee jobs and self-employment. Individuals are quite young with a high education, and they are engaged in medium-high skilled professions. They are mostly Italians from central and northern areas, and women are relatively more present. This cluster has much in common with cluster 2, where household services (mainly recreation and health) have an appreciable specialisation: in this case, high education is combined with high-skilled employee jobs and older individuals.

Concluding remarks

The availability of household survey microdata is essential for a deeper understanding of the underground employment complexity: the main challenge of this approach is represented by the micro-level indirect detection of irregular job holders and by the correction of the under-coverage induced by this latent phenomenon. The statistical integration of survey microdata with administrative records is a promising approach for the building of an integrated large survey sample when the flagging of irregular jobs and employment is combined with the treatment of biases.

In this work we have tested the possibilities offered by the integrated LFS-ADMIN sample developed at Istat to support national account benchmark estimates. The descriptive analyses of this data set confirm the main results of previous research, adding more details on irregular employment, especially with reference to the heterogeneous characteristics of individuals and of their economic environment. A measurement of their effect on the probability of working underground highlights the relevance of factors connected with weaker individual positions in the labour market. These factors depend strongly on local conditions, so that the same individual profile may be characterised by quite different probabilities of being underground according to whether the local labour markets are or are not endowed with appreciable inclusion capabilities: high inactivity rates, large *grey areas* and scarcity of active policies are all clearly associated with higher probability of irregularity. The economic environment plays an important role: low tax compliance and a higher weight of very small firms offer a larger room for underground work.

The segmentation of irregular employment shows how heterogeneous is the combination of labour supply conditions with actual labour demand. This evidence gives the possibility to appreciate

the coexistence of different irregularity profiles obtained by combining sector and socio-economic conditions. Such results would suggest the adoption of an approach to active labour market policies where local conditions should receive major attention.

Future research involves in the near future a refinement in the shaping of the LFS-ADMIN sample through the enlargement of the set of ADMIN sources. In the medium term, developments should be aimed at a more efficient use of ADMIN data to improve the breakdown of estimates, for instance through approaches based on small area estimations. A longer term challenge could be the analysis of the interactions between regular and irregular side of the labour market at local level.

References

- AA.VV. *Soluzioni metodologiche per l'utilizzo integrato delle fonti statistiche per le stime dell'occupazione*. Istat working paper (in referaggio). Roma: Istat, 2014
- Baccini A., L. Castellucci e M. Vasta. *Lavoro sommerso: il caso della Toscana*. Studi e note di economia n. 1, Firenze: 2003
- Boeri T., Garibaldi P. *Shadow activity and unemployment in a depressed labour market*. CEPR discussion paper n.3433. Roma: CEPR, 2002.
- Calzaroni M. *L'occupazione come strumento per la stima esaustiva del PIL e la misura del sommerso*. (Atti del seminario "La nuova contabilità nazionale", 12-13 gennaio). Roma: Istat, 2000
- Cappariello R., Zizza R. *Dropping the books and working off the books*. Temi di discussione n. 702. Roma: Banca d'Italia, 2009
- De Gregorio C., Filippini D., Martini A., Rocchetti I. *A comparison of sample and register based survey: the case of labour market data*. (Paper presented at Q2014 Conference). Vienna: Statistik Austria, 2014
- De Gregorio C., Giordano A. "Nero a metà": *contratti part-time e posizioni full-time fra i dipendenti delle imprese italiane*. Istat working paper, n.3. Roma: Istat, 2014
- García Martínez M.A., Galindo J.O. *Integrating administrative data into the LFS data collection. The Spanish experience obtaining the variable INDECIL from administrative sources* (INE working papers n.8). Madrid: INE, 2011
- Hochfellner D. *Dealing with inconsistencies in the integration of multiple register based data sources*. (NTTS paper). Luxembourg: Eurostat, 2011
- Iso. *La terziarizzazione del sommerso dimensioni e caratteristiche del lavoro nero e irregolare nel settore dei servizi*. I libri del Fondo sociale europeo, n.154. Roma: Iso, 2007a
- Kuijvenhoven L., Scholtus S. *Bootstrapping Combined Estimators based on Register and Sample Survey Data*. (Statistics Netherland discussion paper, n.23). Den Haag: CBS, 2011
- Linder F., van Roon D., Bakker B. *Combining data from administrative sources and sample surveys; the single-variable case. Case study: Educational Attainment*. In: ISTAT, CBS, GUS, INE, SSB, SFSO, EUROSTAT, ESSnet on Data Integration. Report on WP4. Case studies. Luxembourg: Eurostat, 2011
- MEF. *Economia non osservata e flussi finanziari*. (Rapporto finale sull'attività del gruppo di lavoro sull'"Economia Sommersa e Flussi Finanziari"). Roma: Ministero dell'Economia e delle Finanze, 2011
- OECD. *Measuring the non-Observed Economy. A Handbook*. Paris: OECD, 2002
- Williams C.C. *Beyond the formal/informal jobs divide: evaluating the prevalence of hybrid 'under-declared' employment in South- Eastern Europe*. (International Journal of Human Resources and Management, n. 21, p.2529-46). 2010
- Zhang L.-C. *Topics of statistical theory for register-based statistics and data integration*. (Statistica Neerlandica, vol. 66, nr. 1, pp.41-63). Den Haag: CBS, 2012
- Zizza R. *Metodologie di stima dell'economia sommersa: un'applicazione al caso italiano*. Temi di discussione n. 463. Roma: Banca d'Italia, 2002