istat
working
papers

N. 1
2015

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

Carlo De Gregorio, Annelisa Giordano

Istat
Istituto Nazionale
di Statistica

istat working papers

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

N. 1
2015

Carlo De Gregorio, Annelisa Giordano

Istat

Comitato scientifico

Giorgio Alleva, Emanuele Baldacci, Francesco Billari
Tommaso Di Fonzo, Andrea Mancini, Roberto Monducci
Fabrizio Onida, Linda Laura Sabbadini, Antonio Schizzerotto

Comitato di redazione

Alessandro Brunetti, Patrizia Cacioli, Marco Fortini
Romina Fraboni, Stefania Rossetti, Daniela Rossi
Maria Pia Sorvillo

Segreteria tecnica

Daniela De Luca, Laura Peci, Marinella Pepe, Gilda Sonetti

Istat Working Papers

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

N. 1/2015

ISBN 978-88-458-1828-8

© 2015
Istituto nazionale di statistica
Via Cesare Balbo, 16 – Roma

Salvo diversa indicazione la riproduzione è libera,
a condizione che venga citata la fonte.

Immagini, loghi (compreso il logo dell'Istat),
marchi registrati e altri contenuti di proprietà di terzi
appartengono ai rispettivi proprietari e
non possono essere riprodotti senza il loro consenso.

The heterogeneity of irregular employment in Italy: some evidence from the Labour force survey integrated with administrative data[1]

Carlo De Gregorio, Annelisa Giordano[2]

Abstract

The intrinsic heterogeneity of irregular employment is analysed by exploiting microdata derived from the statistical integration of the Labour force survey sample with administrative records tracing regular jobs, whereby irregularity is flagged by comparing independent sources. Following previous approaches, logistic regression is used to model the probability of being an irregular worker as a function of individual characteristics and local context indicators. A segmentation of irregular employment shows how the combination of labour supply conditions with actual labour demand is heterogeneous. The results obtained give the possibility to appreciate the coexistence of different specialisation patterns deriving from the combination of sector-related 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 analysis, Cluster analysis.

Sommario

La natura profondamente eterogenea dell'occupazione non regolare viene qui analizzata attraverso l'uso del campione della Rilevazione sulle forze lavoro i cui microdati sono stati integrati con le informazioni contenuti negli archivi amministrativi che tracciano l'occupazione regolare. Attraverso una regressione logistica 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 alcuni modelli di specializzazione attraverso la combinazione fra caratteristiche dell'offerta e struttura della domanda, evidenziando la coesistenza di profili eterogenei.

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

[1] A preliminary version of the present paper was submitted on the occasion of Istat event Giornate della Ricerca on 10-11 November 2014.

[2] cadegreg@istat.it, giordano@istat.it. Any opinions expressed in this working paper are those of the authors and not those of ISTAT.

Index

Pag.

Introduction 6
1. The LFS-ADMIN integrated sample and the identification of irregular workers 7
2. Modelling the probability of working underground 10
3. A segmentation of irregular employment..... 14
4. Concluding remarks 16
References 17
Appendix 19

Introduction

Research on irregular employment[3] has progressively moved from the mere estimation of general indicators concerning the incidence of this phenomenon towards more explicit insights into its multi-facet nature, contributing to enrich analysis and – potentially – to address and support suitable and dedicated policies[4]. Heterogeneity of hidden employment reasonably derives from the interaction of individual characteristics (such as age, gender, education, skill, household structure, economic conditions, etc.) with those of the surrounding economic and social environment (as summarised, for instance, by the functioning of local labour market and active labour market policies, by income distribution, by the structure of the business sector, and by the attitudes towards tax compliance). Further sources of heterogeneity derive from the definition of the boundaries of irregularity and in particular from the increasingly fading borders between regular and irregular labour input: as a matter of fact, irregular labour input actually derives not only from straight irregular jobs (primary or secondary) but also from formally regular jobs, due to unreported working time with partial evasion of social security and tax duties. The importance of this so called grey economy might also be envisaged as the result of a partial adaptation to policy actions tailored to contrast purely hidden jobs[5]. As a consequence, accurate estimates of grey labour input should necessarily be based on actual working time: at the same time, it is increasingly embarrassing to represent irregularity as a headcount binary variable, while continuous or k-way categorical variables would better satisfy this purpose.

Research on this issues is still on the way, and the statistical integration of survey and administrative data looks like a very promising path towards the provision of helpful insights on hidden labour input, consistently with accurate level estimates. In what follows the focus is on irregular employment *tout court*; a source integration approach was adopted with the objective of outlining a hypothesis of segmentation of purely hidden labour market.

The use of microdata is fundamental for this purpose. In the recent past important achievements have been obtained by Istat in estimating irregular labour input by means of aggregated (or *macro*) approaches, mainly founded on the cross comparison of detailed domain aggregations of employ-

[3] 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).

[4] 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.

[5] 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. Boeri *et al.* (2002) focused instead on the fading borders between irregularity and unemployment.

ment data from independent sources[6]. This approach guaranteed as a matter of fact accurate level estimates of irregular labour input in Italy, with an appreciable breakdown at least for the national accounts purposes it was meant to satisfy: nevertheless, macro approaches are not suited for the provision of detailed analyses of hidden labour market. More recently, the analysis of household survey microdata has gained ground, based on the indirect detection of irregularity at individual level by selecting the answers to groups of items of household survey questionnaire. Cappariello *et al.* (2009)[7], in particular, derive very interesting results by flagging individuals in employment as irregulars if they do not declare social security coverage[8]; Boeri *et al.* (2002) worked on a survey sample limited to Sicily where irregularity was directly asked in the questionnaire. These approaches, if on one side they do not meet the target of providing unbiased level estimates (mainly because they cannot exploit source integration to correct the response biases), on the other side they paved the way for a deeper study of individual characteristics and hidden jobs segmentation.

The paper develops this latter approach by exploiting microdata derived from the statistical integration of the Italian Labour force survey (LFS) sample with administrative records (hereafter summarised with ADMIN) tracing regular jobs, where irregularity is flagged by comparing the employment status reported by independent sources. This integrated sample (named LFS-ADMIN[9]) has the advantage of allowing the use of a huge amount of microdata where the detection of irregularity is derived within a statistical integration process that corrects employment level bias[10]. In what follows a short presentation of LFS-ADMIN with a description of irregular work estimates is given (par. 1); some results obtained from modelling the probability of being in irregular employment are discussed (par. 2); then a segmentation of irregular employment based on individual and job characteristics, as well as on the ADMIN traces of each individual, is provided (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 more than 600 thousands interviews representative of individuals in the resident population[11]. LFS-ADMIN integration has brought to the estimate of the actual employment status of each individual in the sample based on a statistical model aimed at reconciling the information for a same individual independently gathered through LFS and ADMIN. In fact, if on one side LFS individual status is referred to a particular week in the year (the reference week), on the other side ADMIN sources have differing levels of precision in detailing the dates of the actual labour input, usually very accurate for employees and more vague for

[6] See for example Calzaroni (2000) whose approach has been founded on the comparison between Census data and Labour force survey data, and Baldassarini (2001). 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).

[7] They analyse microdata of the biennial Survey on Household Income and Wealth, run by the Bank of Italy with a sample of nearly 8.000 households.

[8] 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.

[9] 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 of private enterprises in industry and services (INPS-EMENS), in recreation (ENPALS), agricolture (INPS-DMAG), of households as employers (INPS-Lavoratori domestici), of public administrations(INPDAP) and on self-employed such as collaborators (INPS-Gestione separata and INPS-Collaboratori professionali), owners in the business sectors (Sistema informativo ASIA-Indipendenti), and in agricolture (INPS-Autonomi agricoli). All these sources have been used as input to build the employment register (DB Occupazione) supporting ISTAT system of business registers (ASIA). A first experience at Istat on survey and ADMIN sources integration is documented in Cascioli (2006).

[10] 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).

[11] Although officially resident, 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 segments is very high. National accounts estimates on the contrary are exhaustive.

self-employment. Furthermore, the definition of the employment status in LFS and ADMIN necessarily differs. LFS adheres to ILO standards: in principle it covers any kind of labour input, regular or irregular. On the contrary, ADMIN status is mostly referred to administrative rules that do not necessarily match ILO standards: for instance, it only refers to labour input with official traces and thus excludes by definition entirely irregular jobs; furthermore it may include false positives.

Tackling reconciliation implies the adoption of methods to detect, measure and correct the biases affecting both sources[12]: namely, the possible under-coverage of employment and particularly of secondary jobs by LFS[13], and ADMIN over-coverage of regular jobs and lack of coverage of irregular work[14].

Irregular jobs have been defined as employment spells unmatched in the reference week with validated ADMIN signals. The integrated dataset lists the jobs performed by individuals actually employed, with indication of the order of the job (primary, secondary, etc.), regularity status, 4-digit Nace, actually worked hours, type of employment, tasks and duties undertaken in the job (coded through Isco), business register data on the employer and the rest of LFS information collected through the survey questionnaire[15]. Integrated job data are thus combined with the personal characteristics of the worker and with the whole profile of his yearly ADMIN records[16]. Since the focus here is on employment, only the individuals in employment according to the integrated estimates[17] have been selected from LFS-ADMIN, by considering only their primary job (be it regular or irregular)[18]: considering both years together, the sub-sample consists of about 480.000 individuals, 48.000 of which with an irregular primary job[19]. LFS weights are used for the grossing up.

LFS-ADMIN estimates for the whole period 2010-2011confirm some expected characteristics of irregular employment already highlighted by other independent estimates[20]. The incidence of underground employment is estimated nearly 10% of total employment in the target population (Table 1). Higher rates can be found among women, foreign citizens (especially from EU countries), self-employed, young people, low education segments, South, and in agriculture, constructions, hotels and restaurants, households services. Other aspects stand out clearly: elderly people seem affected by higher rates, like low skilled professionals; households structure and the role of the individual within the household play a non-secondary role; the presence of other irregular workers in the households is also associated with larger irregularity rates.

[12] The methods adopted for data integration are fully described in AA.VV. (2014) and will not be discussed further here. For a short outline see also 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). Pavlopuolos *et al.* (2012) tackle the issue of the lack of a benchmark between survey and administrative data in the measurement in temporary employment. All these models face data integration as conditional probability estimates. Fuzzy variables techniques could be a very promising tool to measure irregularity, at the moment unexplored.

[13] Boeri et al. (2002), for instance, affirm that a meaningful share of unemployed and inactive LFS respondents are actually employed in the informal sector. See AA.VV. (2014) for a deeper insight of this issue.

[14] ADMIN over-coverage is source dependent. It depends in fact on the available information concerning the accuracy of the dating of actual labour input. A lack of precision affects mainly the sources on self-employment. On the contrary, those on employees are usually very precise and report duration and dates of labour contracts. See AA.VV. (2014) for more details.

[15] 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. Statistical imputation (generally hot-deck donor imputation) is used for the LFS individuals rescued from employment under-coverage. See AA.VV. (2014) for details.

[16] In perspective, ADMIN data can be organized longitudinally and individual regular histories can be used more efficiently to outline and detect irregularity.

[17] They include thus all the people in employment according to LFS plus the remaining individuals rescued from LFS bias thanks to ADMIN signals and undercoverage estimates.

[18] 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.

[19] About 55 thousands secondary jobs have been excluded, 8.000 of which correspond to irregular jobs. All these figures are very similar in 2010 and 2011.

[20] 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).

Table 1. Irregularity rates and irregular employment shares by segment. Two-year 2010-2011 (%)

Segment	Irregularity rate (a)	Share on total irregulars (b)
TOTAL	9.8	100.0
GENDER: Men	9.0	53.9
GENDER: Women	11.0	46.1
MARITAL STATUS: Unmarried	13.0	41.5
MARITAL STATUS: Divorced or widow	11.5	9.1
CITIZENSHIP: EU	21.7	6.4
CITIZENSHIP: Extra EU	18.0	10.9
AGE: 15-24 yrs.	21.3	12.0
AGE: 55-64 yrs.	23.2	3.8
AGE: 65 yrs. or more	36.7	1.0
Isced: Primary education or less (Isced 0&1)	19.7	11.6
Isced: Lower secondary education (Isced 2)	11.1	34.9
HOUSEHOLD: Single	13.7	16.5
HOUSEHOLD: Child, with both parents	14.6	19.4
HOUSEHOLD: Child, with single parent	13.0	5.5
HOUSEHOLD: Presence of irregular job holders	17.9	13.4
NACE: Agricolture	21.6	8.9
NACE: Construction	12.8	10.8
NACE: Hotel and restaurants	16.1	8.4
NACE: Recreation	25.8	3.3
NACE: Other households services	21.5	6.4
NACE: Households as employers	29.8	8.4
Nuts1: South & Islands	15.7	45.5
Nuts2: Campania	19.5	14.5
Nuts2: Calabria	19.8	5.3
Nuts2: Sicilia	15.1	9.9
TYPE OF EMPLOYMENT: Self-employed	12.8	35.3
ISCO: Skilled agricultural workers (ISCO 6)	15.8	3.7
ISCO: Elementary occupations (ISCO 9)	19.7	19.6

Source: LFS-ADMIN, Two-year 2010-2011
(a) Irregular employment as percentage of total employment in the segment.
(b) Irregular employment in the segment as percentage of total irregular employment.

Integration delivers further challenging hints. For instance, the contiguity between regular and irregular jobs: underground workers frequently are traced in ADMIN during the reference year: in other words, such traces are not compatible with any coverage in the reference week of LFS interview but anyway characterize the working activity of the individual in other parts of the year. This seems to indicate a switching from regularity, suggesting again that the treatment of irregularity should be followed through continuous or at least multi-modal variables: it is also worth mentioning that such evidence seems to encourage the adoption of a fuzzy variable approach to target irregularity[21]. It is also worth mentioning that an important share of irregulars declares to LFS to work in large local units. These aspects open the way for analysis dedicated to underground outsourcing of services by larger enterprises. 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. This appears another promising subject for further research. Although the well-known stereotypes of irregularity are evidently confirmed,

[21] This approach has been for instance adopted for the analysis of poverty (see Betti *et al.* 2009). For a general overview of fuzzy variables see, *inter alia*, Colubi *et al.* (2007).

the integrated sample confirms also that heterogeneity lays behind them[22]. Irregularity is spread across many segments of the labour market, although with different intensities, and this basic fact deserves a special focus in order to target the analysis and support policy.

2. Modelling the probability of working underground

The probability for a person in employment to work underground can be analyzed from different perspectives. It was chosen to start from the individual characteristics (such as age, gender, household, citizenship, etc.) which have been primarily used as independent variables in the analysis: other variables have been progressively introduced to summarize local context effects.

A battery of Nuts3[23] labour market indicators is used to monitor the effects induced by actual local market and active policy conditions on individual profiles: it was opted for the rates of activity and unemployment, the contact rate of job centers (be it public or private)[24], the coexistence rates of the so called *grey area* with official labour force and of potential employment[25] with total employment[26]. It is worth noticing that Nuts3 is a planned estimation domain in LFS sample design: for Italian LFS the Nuts3 coefficients of variation for the unemployment rate ranged from 3.7 to 30.8 in 2011[27], with a median of 10.2% and 5%-95% percentiles respectively equal to 5.1-18.8%[28].

A partition of the Nuts3 into eight clusters - derived from the DBGEO database developed by the tax authority[29] - has been used as a proxy of the local attitudes towards tax compliance[30]. The effects of the employment structure of local regular business have been summarized with sector and firm size indicators by gender, all derived from the integrated sample.

Finally, a last set of input variables concerning the actual job of the individual - Nace and type of employment (employee or self-employed) - is used to introduce the demand side of the irregular labour market.

By using the nature of employment - whether regular or irregular - as the response variable, a logistic regression has been run to model the probability of this event in function of the above mentioned sets of variables. Several specifications have been tried, changing the sets of variables, the interactions and the model groups[31]. What follows is the general simple effect version:

[22] Cappariello *et al.* (2009) also stress this point.

[23] Nuts3 classification level corresponds to the more than one hundred "province" in which Italy is split. This level is actually an estimation domain in LFS sampling design. Lower levels of territorial disaggregation, such as Lau1 (corresponding to Nuts4) and Lau2 (Nuts5), have not been considered here.

[24] 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 that a higher use of official channels is an indicator of active policy concern and marks an antibody against informal jobs.

[25] As defined by the sum of unemployed and *grey area*.

[26] 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, their first three principal components, estimated by gender, were also used. 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.

[27] Eurostat (2013), ch.9.

[28] These latter data are derived from Istat Information System on Quality (SIQual, http://www.istat.it/en/tools/data-quality) and are referred to 2006 data.

[29] 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.

[30] A first cluster, called *the Equilibrist*, groups small Nuts4 with medium living standard and tax compliance; *the Industrial* gathers industrial territories relatively compliant; *Metropolis* are the urban areas with medium-high tax evasion; *Nothing to declare* are small Nuts4 with tax non-compliance and low wealth; *Not angels* are areas with critical compliance and medium-low living standard; *Risky habits* are weak local economies, with criminality and medium compliance; *Total risk* characterized by very low compliance and very low living standards.

[31] 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.

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

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 primary job of each individual.

Some results obtained with three simple effects models run separately on the two gender groups are reported hereafter: 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 from model A to model C: for instance, for both model groups the concordance ratio ranges from about 67-68% to 73-74%[32]. Both groups show that foreign citizens have a higher probability of being in irregular employment: within this segment, 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 the ratio drastically decreases - although only for males - remarking the importance of the demand side factors. Age appears characterized by some symmetry: the probability of being irregular grows as the distance from central age classes increases, especially for ancient males. It's worth noticing that for young people age and household effects add up, given the higher odds associated to individuals living with parents. Some differences between the genders emerge considering 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[33]. Another class with a higher irregularity risk is the class of single parents living with sons; here the odds nearly double the benchmark. Household income is also important in determining the risk of irregularity: the presence of another income works quite differently according to whether it is regular (slightly lower risk) or irregular (much higher risk). A low education attainment is confirmed to be a crucially risky condition, even harder for women. 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)[34].

Labour market conditions seem to operate differently by gender. The risk of irregularity for men increases 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 irregular side of the market. This does not seems to be the case for men, for whom higher contact rates with job centers seem to reduce the risk of irregularity. DBGEO clustering copes well with explaining underground work for both genders, in particular when included in the *Total risk* cluster. The effect associated with the structure of regular business deserves some attention. For men, the higher the relative weight (in terms of regular employees) of difficult sectors such as agriculture and construction the lower the probability of underground jobs: this might be connected to the emersion of previously underground activities[35]. On the contrary, the relative weight of regular employees in household services and in microenterprises seems related to higher irregularity risks.

The introduction into the analysis of details on irregular jobs brings into light other gender dif-

[32] See the tables A.1-A.3 in the Appendix for details on model fit and estimates.

[33] It should be noticed that the household here described derives from population registers, and might not coincide with the actual "economic household".

[34] This aspect, however, needs further analysis in order to explain why the same is not 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.

[35] This effect is anyway not at work for women.

ferences. 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.

Considering as benchmark one of the most virtuous profiles (profile 1: a middle-aged highly educated male in a household with a regularly employed wife and at least two sons. See Table 2), model A predicts for him 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 a 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.

Profile 2 describes a young individual living with both parents in a household with a least one regular income: his predicted probability goes near to 20% and near 30% if she was a girl. Both probabilities are over respectively 30% and 40% if the household income was irregular. A higher Isced would reduce both probabilities and the gender distance. But if profile 2 was an EU citizen 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. Profile 5 describes what happens to the son if such a parent is not an income earner: if male, his prediction would be 26%, 37% if his parent was an irregular himself and respectively 32% and 45% if female.

The adoption of model B and C introduce variability in these profiles. The prediction for male in profile 1 ranges from 2% to 11% if context factors are introduced and its maximum peaks 19% with model C predictions; for women the right tail of the distribution is prolonged. 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 context factors are accounted for, and may pass 70% if the type of job 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.

It is worth the while to remind that these results are derived from a sample, though a quite large one, which is representative of resident population. This means that while they can be easily generalized to the observed population, in order to fully accomplish this purpose it is still needed an accurate definition of the sampling error associated to the statistical integration of the different sources[36]. Furthermore, it must be kept firmly into consideration the fact that the observed population does not include those foreign citizens whose presence in Italian territory is not regular. Those people are by definition also irregular workers, but their structural characteristics are rather peculiar in terms of age, citizenship, gender, skill, education even if compared with those of the foreigners who are instead regularly present in Italy. For this reason, our results can hardly be generalized to this segment of the present population.

[36] De Gregorio, Filipponi *et al.* (2014) moved some steps forward in this direction following the developments of the ESSnet on Data integration. See also García Martínez (2011), Hochfellner (2011), Kuijvenhoven et al. (2011), Linder *et al.* (2012), Pavlopuolos *et al.* (2012), Zhang (2012). It has to be mentioned that replication techniques and bootstrapping have been used in order to validate these estimates of irregular labour input for national accounts purposes: they- provided encouraging results.

Table 2. Predicted probability of being a person in irregular employment, by citizenship, profile and model *(%)*

					MALE					FEMALE				
						B		C			B		C	
Profile	Age	Role and Hh structure	Isced	Other incomes	A	min	max	min	max	A	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
1.1	35-54	Spouse (2 Parents & ≥2 sons)	2	REG	5.5	2.6	14.3	1.9	22.9	10.0	4.5	22.2	2.8	40.5
1.2	35-54	Spouse (2 Parents & ≥2 sons)	2	none	7.2	2.7	14.7	2.0	23.6	12.9	4.8	23.6	3.7	41.0
1.3	35-54	Single	2	none	12.1	5.9	27.9	4.4	40.7	12.7	5.6	26.7	3.8	47.2
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
2.1	15-24	Son (2 Parents & ≥2 sons)	2	IRREG	35.3	15.3	52.0	11.6	63.5	48.6	32.2	62.7	26.4	73.9
2.2	15-24	Son (2 Parents & ≥2 sons)	5	REG	15.3	7.5	33.6	6.1	45.4	20.8	9.7	39.6	7.6	66.5
2.3	25-34	Son (2 Parents & ≥2 sons)	2	REG	13.4	6.3	29.4	4.9	38.9	19.4	10.2	36.7	7.2	59.7
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
4.1	35-54	Parent (1 Parent & ≥2 sons)	3-4	none	5.4	2.8	12.4	2.5	16.8	8.0	3.8	17.8	2.5	25.7
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
5.1	15-24	Son (1 Parent & ≥2 sons)	2	IRREG	37.3	17.3	52.8	14.2	58.0	44.6	27.7	58.1	21.9	63.1
									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
1.3	35-54	Single	2	none	26.8	16.2	52.9	11.9	60.2	27.3	16.8	48.8	11.3	63.3
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
4	35-54	Parent (1 Parent & ≥2 sons)	2	none						27.9	18.1	40.4	12.7	41.8
									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
1.3	35-54	Single	2	none	19.7	13.2	48.6	10.4	61.2	20.9	11.5	42.8	9.0	60.5
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
4	35-54	Parent (1 Parent & ≥2 sons)	2	none						21.4	12.8	34.9	9.8	41.5

Source: LFS-ADMIN, years 2010-2011

3. A segmentation of irregular employment

The individuals in LFS-ADMIN with an irregular primary job have been analysed by means of a sequential use of correspondence analysis (MCA) and Ward hierarchical clustering[37]: 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 ADMIN traces have been added to *W*, scaled according to the intensity of ADMIN signals[38].

With nearly 50 variables and 150 modalities the first ten eigenvalues accounted for about 40% of total inertia[39]. Better results were obtained by replying separate sector analyses, although the structure of the data base, as revealed by the first components, appears relatively stable if MCA is separately run by economic sectors. The results from the overall sample are reported hereafter. In general, the first component (6.5% of total inertia; see Chart A.1 and Chart A.2 in the appendix) opposes two poles that could be summarized as "*unskilled blue collars*" vs. "*skilled self-employed*"[40]. The second component (5.8%) offers a further distinction somehow specular as compared to the first one: "*skilled white collars*" vs. "*low education self-employees*"[41]. The third component (4.6%) opposes the activities of "*foreign women*" vs. "*aged & skilled craft workers*".

Ward clustering on the first ten MCA components delivers a tentative classification of irregular employment: the description of a nine cluster partition is reported in Table 3. Such partition explains more than 64% of total inertia. Construction and household services show a specialisation in cluster 3 (the largest), characterised by low skill part-time employee jobs: it sounds reasonable to find in this cluster a relatively strong presence of residents in EU countries. 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 workers from central and northern regions are relatively more frequent. Industry itself is strongly present in cluster 9 where employees have intermediate skill levels and are more frequently partially traced in the annual ADMIN: northern regions and EU residents have some ties with this group.

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

[37] Fuzzy clustering could eventually be applied in order to take account of the fuzziness of the concept of irregularity.

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

[39] Given the large number of variables and modalities, and as a consequence given the high number of eigenvalues of MCA, the share of explained inertia is relatively appreciable. Low explained inertia does not mean that the analysis is not valid, but it does mean that extra care should be eventually taken in interpreting the plot. A reevaluation technique might be applicable anyway.

[40] 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.

[41] 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 3. Clusters of individuals in irregular employment, by Nace, gender, age and citizenship *(distribution; specialization rates by segment; %)*

			Specialization rates											
				Nace										
Cluster	Short description	Distr.	Total	Agriculture	Industry	Construction	Trade & Horeca	Business services	Household services	Women	Young 15-34	EU	Extra EU	South
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	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	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	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	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	Employees, Elementary occupations, Low education, South, Parent, 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	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	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	Blue collars, traces in ADMIN, Some 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, Two-year 2010-2011

Cluster 5 shows a meaningful presence of persons employed in the business and household services: this cluster appears 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 services) have an appreciable specialisation: in this case, high education is combined with high skill employee jobs and older individuals.

The connection between the individual characteristics of irregularity with the individual traces present in ADMIN sources (that belong to the regular side of the market) suggests a deeper scrutiny. On one side, the flows from regularity to irregularity (and vice versa) can be deemed as strongly dependent on the nature of individual labour market "stories" (quality, experience, age). On the other the patterns of irregularity look somehow ADMIN-dependent in the sense that they seem to have adapted to sector specific habits and needs and to and regional influences.

Concluding remarks

The availability of household survey microdata is essential for disentangling the complexity of underground employment. The main challenge of this approach is represented by the micro-level indirect detection of irregular job holders and the correction of the under-coverage associated with the latent nature of the phenomenon. The statistical integration of large survey sample microdata with administrative records is a promising approach since the individual flagging of irregularity can be combined with the treatment of employment status biases.

In this work the possibilities offered by the integrated LFS-ADMIN sample developed by Istat to support national account benchmark estimates have been tested. The descriptive analyses of this data seem to confirm the results derived from previous research on this subject adding more details on irregular employment, especially with reference to the heterogeneous characteristics of individuals and of their environment. A measurement of the effect those characteristics on the probability of working underground highlights the relevance of factors that appear connected with weaker individual positions in the labour market. These factors seem strongly dependent 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, large *grey areas* and scarcity of efficient policy actions are all presumably associated with a higher probability of being irregular. It seems that the local economic environment actually plays an important role: low tax compliance and a higher weight of very small firms offer larger room for underground work. This aspect and the causal links have to be further investigated with a dedicated methodological approach.

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 models of irregularity obtained by combining sector and socio-economic conditions that reveal quite reasonable specialisation patterns. Such results would suggest the need to adopt coordinated approach to contrast irregularity, based on active policies and where local conditions should receive greater attention. Quite evidently also this suggestion need confirmation.

Though encouraging, our results also deserve some further deepening under several profiles that concern both definitional and methodological issues. The boundaries of irregularity need in particular to be accurately scrutinised. In our approach, irregular employment corresponds to working activities not traced in any administrative register: that may include also activities that simply are not subject to any administrative obligation (as for example it may happen for very small scale self-employment in agriculture). Furthermore, implicit in LFS-ADMIN there is the idea that LFS records only perfectly legal businesses, although possibly illegal with respect to social security and tax compliance: this assumption needs to be verified, with the help of the advancement that are taking place in the measurement of illegal economy. The most important definitional issue has anyway to

do with the need to fully consider the grey economy within the context of irregularity analysis. This aspect seems extremely relevant under the economic point of view: its measurement involves progresses in the estimates of working time both from households and business statistics sides[42].

The methodological aspects are anyway those who appear more promising, both for data integration in LFS-ADMIN and for the analysis of irregularity. The approach actually based on logistic regression might exploit further advancements in this area and in particular moving from traditional statistical analysis to causal analysis of multivariate data in particular for the evaluation of the efficacy of labour market active policies. For the same purpose, propensity score matching approaches are worth to be tested. The use of more sophisticated approaches based on logistic regression is also worth to be tested in order to face more properly the events associated with measurement error in covariates.

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 quite challenging research activity, starting from LFS-ADMIN, could be oriented to the analysis of the interactions between regular and irregular side of the labour market at local level. Finally, an entirely new approach would consider the idea of turning upside-down the logic behind LFS-ADMIN: passing from the integrated LFS-ADMIN sample to the exploitation of LFS-ADMIN inference in order to fully exploit the information in the whole set of ADMIN data, which cover the universe of the formally regular jobs the present population is engaged in.

References

AA.VV. *Soluzioni metodologiche per l'utilizzo integrato delle fonti statistiche per le stime dell'occupazione.* Final report of Istat Working group on the Development of new methods for employment estimates for National accounts benchmark. Mimeo, to be submitted for publication as Istat working paper. 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

Baldassarini A. *Non regular foreign input of labour in the new national accounts estimates*. OECD meeting of national accounts experts. Paris, 9-12 October. Paris: OECD, 2001

Baldi C., De Gregorio C., Giordano A., Pacini S., Solari F., Sorrentino M. *Joint use of survey and administrative sources to estimate the hours actually worked in the business sector*. 1st Southern European Conference on Survey Methodology (SESM). Barcelona: 2013

Betti G., Ferretti C., Gagliardi F., Lemmi A., Verma V. *Proposal for new multidimensional and fuzzy measures of poverty and inequality at national and regional level.* Dipartimento di Metodi Quantitativi, Università di Siena, Working Paper n. 83, November. Siena: 2009

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 seminari*o "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

Cascioli R. *Integrazione dei dati micro dalla Rilevazione delle Forze di Lavoro e dagli archivi amministrativi INPS: risultati di una sperimentazione sui dati campione di quattro province*. Contributi Istat, no. 6. Roma: Istat, 2006

[42] Baldi *et al.* (2013).

Colubi A., Coppi R., D'Urso P., Gil M.A. *Statistics with fuzzy random variables*. Metron International Journal of Statistics vol. LXV, n. 3, pp. 277-303. Roma: 2007

De Gregorio C., Filipponi 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

Eurostat. *Quality report of the European Union Labour Force Survey 2011*. Luxembourg: Eurostat, 2013

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

Isfol. *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: Isfol, 2007a

Isfol. *Le politiche per l'emersione del lavoro nero e irregolare*. *Quadro istituzionale e normativo* (Monografico n. 11 dell' Osservatorio Istituzionale e Normativo). Roma: Isfol, 2007b

Isfol. *Dimensioni e caratteristiche del lavoro sommerso/irregolare in agricoltura*. Area "Si-stemi Locali e Integrazione delle Politiche". Roma: Isfol, 2011

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

Pavlopoulos D., Vermunt J.K. *Measuring temporary employment. Do survey or register data tell the truth?* (EALE 2012 conference). Bonn: 2012

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 Neederlandica, 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

Appendix

Table A.1. Model fit statistics of the logistic regression, by model and model group

	MALE				FEMALE			
Model Fit Statistics	No Inter-cept	Model A	Model B	Model C	No Inter-cept	Model A	Model B	Model C
AIC	164,651	154,984	149,197	147,402	137,044	129,536	125,611	122,890
SC	164,662	155,245	149,626	147,893	137,044	129,789	126,026	123,366
-2 Log L	164,649	154,934	149,115	147,308	137,044	129,486	125,529	122,796
Chi-square test (a)								
Likelihood Ratio		9,716	15,534	17,342		7,557	11,514	14,246
Score		11,082	17,161	19,003		8,841	12,985	15,675
Wald		9,789	14,590	15,935		7,722	11,026	13,090
Other model fit statistics								
Percent Concordant		67.0	72.2	73.4		65.9	70.5	72.8
Percent Discordant		31	26.9	25.9		32.2	28.7	26.5
Percent Tied		2	0.8	0.8		1.9	0.8	0.7
Pairs		5.85E+09	5.85E+09	5.85E+09		3.68E+09	3.68E+09	3.68E+09
Somers' D		0.36	0.453	0.475		0.337	0.417	0.462
Gamma		0.367	0.457	0.478		0.343	0.42	0.466
Tau-a		0.063	0.079	0.083		0.071	0.088	0.098
C		0.68	0.726	0.737		0.668	0.709	0.731

(a) All test statistics have a probability less than 0.0001. The degrees of freedom are 24 for model A, 41 for model B and 47 for model C.

Table A.2. Test statistics for logistic regression variables, by model and model group

		MALE			FEMALE		
Effect	DF	Model A	Model B	Model C	Model A	Model B	Model C
CITIZENSHIP	2	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
AGE CLASS	5	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
HOUSEHOLD STRUCTURE AND ROLE	11	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
EDUCATION	4	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
OTHER HOUSEHOLD REGULAR INCOMES	1	<.0001	0.060	0.016	<.0001	0.000	<.0001
OTHER HOUSEHOLD IRREGULAR INCOMES	1	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
LABOUR MARKET (by Gender) Fact.1 (a)	1		<.0001	<.0001		<.0001	<.0001
LABOUR MARKET (by Gender) Fact.2 (b)	1		0.501	0.514		0.043	0.025
LABOUR MARKET (by Gender) Fact.3 (c)	1		0.000	0.000		0.999	0.690
DBGEO PARTITION	7		<.0001	<.0001		<.0001	<.0001
REGULAR EMPLOYMENT STRUCTURE (by Gender) Agricolture	1		0.613	<.0001		<.0001	<.0001
REGULAR EMPLOYMENT STRUCTURE (by Gender) Construction	1		0.005	0.000		0.218	0.133
REGULAR EMPLOYMENT STRUCTURE (by Gender) Trade	1		0.551	0.844		<.0001	<.0001
REGULAR EMPLOYMENT STRUCTURE (by Gender) Business services	1		0.061	0.204		<.0001	<.0001
REGULAR EMPLOYMENT STRUCTURE (by Gender) Household services	1		<.0001	<.0001		<.0001	<.0001
SHARE OF EMPLOYMENT IN MICROENTERPRISES	1		0.000	0.001		0.525	0.891
TYPE OF JOB (Employee/Self-employed)	1			0.000			<.0001
IRREGULAR JOB NACE CODE	5			<.0001			<.0001

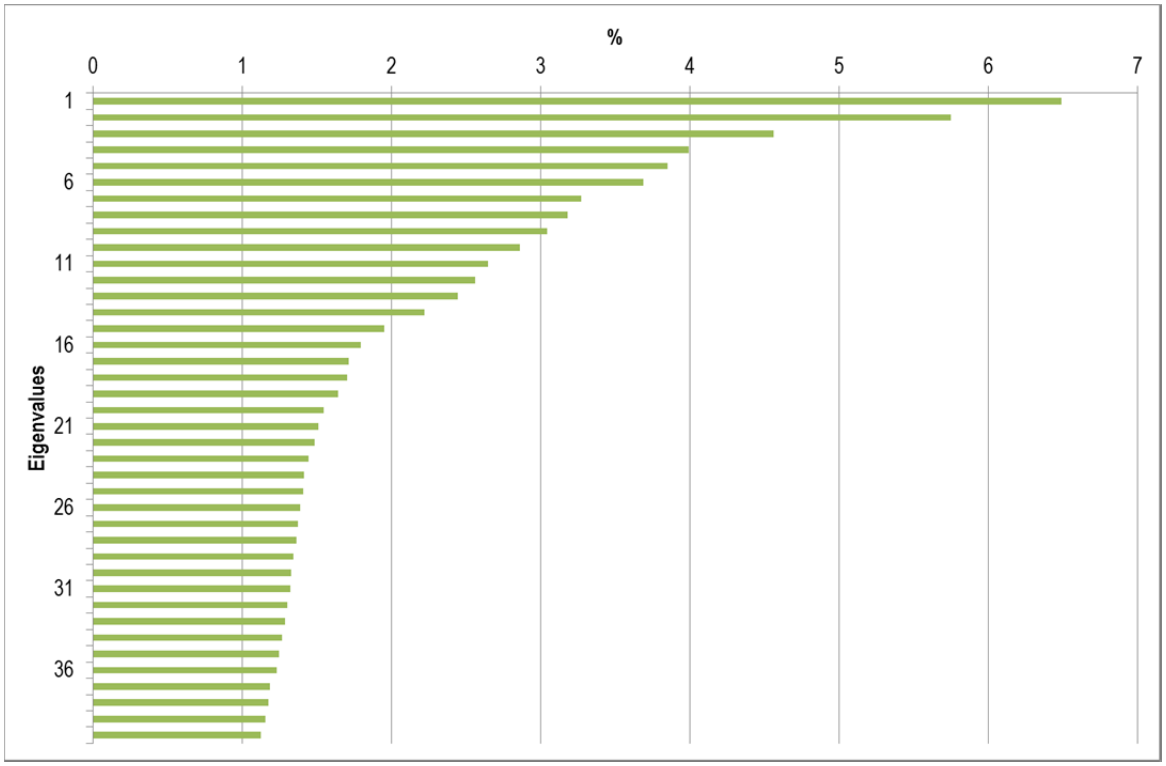
(a) Unemployment, inactivity and grey area vs. virtuous labour market.
(b) Unempoloyment and placement vs. grey area.
(c) Placement.

Table A.3. Parameter estimates, by model and model group

	MALE									FEMALE								
	Model A			Model B			Model C			Model A			Model B			Model C		
Variables and modalities	Est.	St.Err.	Pr.	Est.	St.Err.	Pr.	Est.	St.Err.	Pr.	Est.	St.Err.	Pr.	Est.	St.Err.	Pr.	Est.	St.Err.	Pr.
Intercept	-0.961	(0.0228)	***	-1.458	(0.1206)	***	-1.182	(0.1217)	***	-0.605	(0.027)	***	-2.006	(0.1659)	***	-1.743	(0.1686)	***
CITIZENSHIP=Italian	-0.519	(0.0149)	***	-0.675	(0.0155)	***	-0.644	(0.0158)	***	-0.517	(0.0146)	***	-0.596	(0.0151)	***	-0.607	(0.0155)	***
CITIZENSHIP=EU	0.459	(0.0236)	***	0.458	(0.0242)	***	0.398	(0.0245)	***	0.433	(0.0213)	***	0.414	(0.0218)	***	0.413	(0.022)	***
AGE=15-24	-0.068	(0.0267)	*	-0.119	(0.0271)	***	-0.057	(0.0278)	*	0.098	(0.0333)	**	0.103	(0.0338)	**	0.327	(0.0345)	***
AGE=25-34	-0.492	(0.0188)	***	-0.583	(0.0192)	***	-0.517	(0.0196)	***	-0.467	(0.0252)	***	-0.513	(0.0255)	***	-0.346	(0.026)	***
AGE=35-54	-0.716	(0.016)	***	-0.748	(0.0162)	***	-0.703	(0.0165)	***	-0.748	(0.0232)	***	-0.797	(0.0235)	***	-0.699	(0.024)	***
AGE=55-64	-0.363	(0.0195)	***	-0.451	(0.0199)	***	-0.475	(0.0201)	***	-0.450	(0.0262)	***	-0.541	(0.0266)	***	-0.535	(0.0271)	***
AGE=65-74	0.534	(0.0282)	***	0.621	(0.0286)	***	0.563	(0.0293)	***	0.417	(0.0436)	***	0.468	(0.0441)	***	0.232	(0.0448)	***
HOUSEHOLD=Single	0.221	(0.0235)	***	0.333	(0.0241)	***	0.317	(0.0243)	***	-0.088	(0.0269)	**	0.024	(0.0274)	n.s.	0.007	(0.0277)	n.s.
HOUSEHOLD=Spouse(2Parents&1son)	-0.434	(0.0216)	***	-0.444	(0.0219)	***	-0.431	(0.022)	***	-0.214	(0.0205)	***	-0.194	(0.0207)	***	-0.198	(0.0209)	***
HOUSEHOLD=Spouse(2Parents&≥2sons)	-0.356	(0.0196)	***	-0.481	(0.02)	***	-0.482	(0.0201)	***	-0.066	(0.0189)	**	-0.141	(0.0192)	***	-0.186	(0.0194)	***
HOUSEHOLD=Parent(1Parent&1son)	-0.197	(0.0822)	*	-0.066	(0.0833)	n.s.	-0.085	(0.0837)	n.s.	-0.191	(0.0369)	***	-0.050	(0.0375)	n.s.	-0.026	(0.0378)	n.s.
HOUSEHOLD=Parent(1Parent&≥2sons)	-0.039	(0.0884)	n.s.	-0.026	(0.0903)	n.s.	-0.021	(0.0905)	n.s.	-0.059	(0.0364)	n.s.	-0.030	(0.0367)	n.s.	-0.025	(0.0371)	n.s.
HOUSEHOLD=Spouse(2no sons)	-0.323	(0.0233)	***	-0.280	(0.0237)	***	-0.265	(0.0238)	***	-0.146	(0.0218)	***	-0.091	(0.022)	***	-0.111	(0.0223)	***
HOUSEHOLD=Spouse(2with other components)	-0.266	(0.0569)	***	-0.246	(0.0575)	***	-0.264	(0.058)	***	0.039	(0.0598)	n.s.	0.071	(0.0608)	n.s.	0.025	(0.0615)	n.s.
HOUSEHOLD=Son(2Parents&1son)	0.164	(0.0301)	***	0.121	(0.0305)	***	0.131	(0.0306)	***	0.179	(0.0328)	***	0.129	(0.0332)	***	0.146	(0.0336)	***
HOUSEHOLD=Son(2Parents&≥2sons)	0.410	(0.027)	***	0.270	(0.0275)	***	0.285	(0.0276)	***	0.424	(0.0285)	***	0.283	(0.029)	***	0.305	(0.0293)	***
HOUSEHOLD=Son(1Parent&1son)	0.171	(0.0379)	***	0.248	(0.0385)	***	0.256	(0.0387)	***	0.042	(0.0451)	n.s.	0.029	(0.0455)	n.s.	0.050	(0.0459)	n.s.
HOUSEHOLD=Son(1Parent&≥2sons)	0.496	(0.0395)	***	0.414	(0.0402)	***	0.424	(0.0404)	***	0.262	(0.0472)	***	0.179	(0.0477)	**	0.218	(0.0481)	***
EDUCATION=ISCED 0-1	0.575	(0.0179)	***	0.465	(0.0183)	***	0.402	(0.019)	***	0.743	(0.0215)	***	0.625	(0.0219)	***	0.573	(0.0227)	***
EDUCATION=ISCED 2	0.115	(0.0124)	***	0.099	(0.0126)	***	0.109	(0.0129)	***	0.144	(0.0137)	***	0.146	(0.0139)	***	0.158	(0.0142)	***
EDUCATION=ISCED 3-4	-0.498	(0.0241)	***	-0.308	(0.0246)	***	-0.287	(0.0248)	***	-0.395	(0.0238)	***	-0.212	(0.0244)	***	-0.193	(0.0246)	***
EDUCATION=ISCED 5	-0.158	(0.0131)	***	-0.171	(0.0133)	***	-0.118	(0.0135)	***	-0.338	(0.0134)	***	-0.347	(0.0136)	***	-0.289	(0.0138)	***
OTHER HOUSEHOLD INCOMES=Regular	-0.147	(0.00838)	***	-0.016	(0.00866)	n.s.	-0.021	(0.00871)	*	-0.143	(0.0109)	***	-0.039	(0.011)	**	-0.049	(0.0112)	***
OTHER HOUSEHOLD INCOMES=Irregular	0.269	(0.0111)	***	0.230	(0.0113)	***	0.217	(0.0114)	***	0.258	(0.0128)	***	0.214	(0.013)	***	0.197	(0.0132)	***
LABOURMARKET(by gender)=Factor1(a)				0.093	(0.00729)	***	0.093	(0.00733)	***				0.076	(0.00771)	***	0.073	(0.00778)	***
LABOURMARKET(by gender)=Factor2(b)				-0.006	(0.0092)	n.s.	-0.006	(0.00923)	n.s.				0.025	(0.0122)	*	0.028	(0.0123)	*
LABOURMARKET(by gender)=Factor3(c)				-0.055	(0.0147)	**	-0.058	(0.0148)	**				0.000	(0.0179)	n.s.	0.007	(0.018)	n.s.
DBGEO CLUSTER=Equilibrist				0.007	(0.024)	n.s.	-0.001	(0.0241)	n.s.				-0.025	(0.0265)	n.s.	-0.046	(0.0267)	n.s.
DBGEO CLUSTER=Industrial				-0.183	(0.0277)	***	-0.185	(0.0278)	***				-0.149	(0.0286)	***	-0.142	(0.0289)	***
DBGEO CLUSTER=Metropolis				0.031	(0.0399)	n.s.	0.031	(0.0401)	n.s.				0.048	(0.0456)	n.s.	0.067	(0.0461)	n.s.
DBGEO CLUSTER=Nothingtodeclare				0.055	(0.0309)	n.s.	0.054	(0.0311)	n.s.				0.105	(0.0306)	**	0.071	(0.0309)	*
DBGEO CLUSTER=Notangels				0.031	(0.0301)	n.s.	0.035	(0.0303)	n.s.				0.097	(0.0344)	**	0.113	(0.0347)	**
DBGEO CLUSTER=Riskyhabits				-0.053	(0.0324)	n.s.	-0.054	(0.0325)	n.s.				-0.100	(0.0375)	**	-0.119	(0.0378)	**
DBGEO CLUSTER=Totalrisk				0.274	(0.0335)	***	0.281	(0.0337)	***				0.243	(0.0378)	***	0.269	(0.0381)	***
REGULAR EMPL.STRUCT.(by gender)=Agricolture				-0.001	(0.0026)	n.s.	-0.011	(0.00263)	***				0.033	(0.00359)	***	0.023	(0.00366)	***
REGULAR EMPL.STRUCT.(by gender)=Construction				-0.011	(0.00394)	**	-0.014	(0.00396)	**				-0.017	(0.0135)	n.s.	-0.021	(0.0137)	n.s.
REGULAR EMPL.STRUCT.(by gender)=Trade&Horeca				0.002	(0.00342)	n.s.	0.001	(0.00345)	n.s.				0.018	(0.00309)	***	0.018	(0.00313)	***
REGULAR EMPL.STRUCT.(by gender)=Businessservices				0.005	(0.00266)	n.s.	0.003	(0.00268)	n.s.				0.019	(0.00317)	***	0.018	(0.00321)	***
REGULAR EMPL.STRUCT.(by gender)=Householdservices				0.016	(0.00208)	***	0.012	(0.0021)	***				0.013	(0.00203)	***	0.011	(0.00206)	***
EMPLOYMENT IN MICROENTERPRISES				0.007	(0.00198)	**	0.006	(0.00199)	**				0.001	(0.00182)	n.s.	0.000	(0.00184)	n.s.
TYPE OF JOB=Self-employed							-0.030	(0.00788)	**							-0.356	(0.00884)	***
IRREGULAR JOB NACE=Agricolture							0.487	(0.0196)	***							0.515	(0.0263)	***
IRREGULAR JOB NACE=Industry							-0.396	(0.0162)	***							-0.137	(0.0243)	***
IRREGULAR JOB NACE=Construction							0.130	(0.0156)	***							-0.088	(0.0547)	n.s.
IRREGULAR JOB NACE=Trade&Horeca							-0.216	(0.0147)	***							-0.213	(0.0187)	***
IRREGULAR JOB NACE=Businessservices							-0.304	(0.0158)	***							-0.344	(0.0213)	***

Note: *** if Pr. < 0.0001; ** if 0.0001<Pr<0.01; ** if 0.01<Pr<0.05; * if 0.05<Pr<0.1; n.s. otherwise.

Chart A.1. Firts 40 eigenvalues of MCA



Source: LFS-ADMIN Two-year 2010-2011

Chart A.2. Main modalities on the two main factor's space

Source: LFS-ADMIN Two-year 2010-2011