

The heterogeneity of undeclared work in Italy: some results from the statistical integration of survey and administrative sources¹

Carlo De Gregorio², Annelisa Giordano³

Abstract

The heterogeneity of undeclared employment is analysed by exploiting microdata derived from the statistical integration of the labour force survey 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 undeclared as a function of individual characteristics and local context indicators. A segmentation of irregular employment gives the possibility to appreciate the coexistence of different specialization patterns deriving from the combination of supply and demand effects. They seem to support the adoption of an approach to active policies where local conditions should receive greater attention.

Keywords: Labour market, Undeclared work, Non-observed economy, Logistic regression, Multiple correspondence analysis, Cluster analysis.

Sommario

L'eterogeneità 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 contenute 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 posizione lavorativa. Una segmentazione dell'occupazione non regolare evidenzia alcuni modelli di specializzazione attraverso la combinazione fra caratteristiche dell'offerta e struttura della domanda. Dai risultati sembra emergere un quadro favorevole all'adozione di politiche attive su base territoriale.

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

¹ The authors are the sole responsible for the content of the paper which does not involve at any rate ISTAT.

² ISTAT, cadedgreg@istat.it.

³ ISTAT, giordano@istat.it.

Introduction

Research on undeclared work⁴ has progressively moved from the mere estimation of the incidence of this phenomenon, at most broken down by a few main variables, towards more explicit insights into its multi-facet nature, contributing to enrich analysis and – potentially – to address and support suitable and dedicated policies⁵. For this purpose, the statistical integration of survey and administrative data might prove a promising tool for the provision of helpful insights on hidden work segmentation through the use of microdata and within a methodological approach that addresses the issue of producing accurate level estimates.

The use of microdata is fundamental for this purpose. In the recent past ISTAT achieved important results in estimating irregular labour input by means of an aggregated (or *macro*) approach, methodologically founded on the cross comparison of detailed domain aggregations of employment data from independent sources⁶. 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 throughout the selection of groups of response items within the survey questionnaire. Cappariello *et al.* (2009), in particular, derive very interesting results by flagging individuals in employment

⁴ According to European Commission (2007), “*Undeclared work is defined as any paid activities that are lawful as regards their nature but not declared to public authorities, taking into account differences in the regulatory system of Member States. This definition links undeclared work with tax and/or social security fraud and covers diverse activities ranging from informal household services to clandestine work by illegal residents, but excludes criminal activities*” (p.2). This is coherent with OECD (2002), where it is defined as “*Employment concealed by the enterprises choosing not to respect employment regulations or immigration laws by hiring labour off the books*” (p.38). This kind of employment involves the paid production and sale of goods and services that are unregistered and/or hidden from the state in order to avoid taxes, social security payments, and security standards. For all the other respects the production is perfectly legal. Within this definition, and despite the absence of an unambiguous alternative agreed upon at international level, it shall be dealt here with jobs which are not traced in administrative records. In this sense the term undeclared seems to fit well the objectives of the paper. “Irregular work” will be the only term used as a synonym of undeclared work hereafter, despite of its broader meaning and even if in the literature undeclared work is “*also referred to as the informal, hidden, cash, twilight, dual, subterranean, parallel, underground, second, unofficial, or shadow economy, as well as moonlighting*”, see for example Renooy *et al.*, (2004), Williams (2007), ILO (2010) and ILO (2013).

⁵ See for example Williams *et al.* (2004), Renooy *et al.* (2004). With reference to the Italian case, see for example Lucifora (2003) or Gobbi *et al.* (2007) and Cappariello *et al.* (2009) and their literature review on these issues. See also the Italian version of the latter 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. An important stimulus to policy oriented research on undeclared work has been conveyed by the initiatives of the EU Commission. For a recent overview see for instance European Commission (2007, 2014) or else the Commission Staff Working Document “*Impact Assessment Accompanying the document Draft Proposal for a Decision of the European Parliament and of the Council on establishing a European Platform to enhance cooperation in the prevention and deterrence of undeclared work*” (Brussels, 9.4.2014 SWD(2014) 137 final).

⁶ For an overview of “macro approaches” see the one provided in GHK *et al.* (2009). On the so called “Italian approach” see Calzaroni (2000) whose methodology was founded on the comparison between Census data and Labour force survey data: on the same subject see also 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 of this literature. See also Cappariello *et al.* (2009) or, on a more specific perspective, Baccini *et al.* (2003), Isfio (2007a, 2011).

as irregulars if they do not declare social security coverage⁷; Boeri *et al.* (2002) worked on a survey sample limited to Sicily where irregularity was directly asked in the questionnaire. These approaches, if on the one hand they do not meet the target of providing unbiased level estimates (mainly because they cannot correct the response biases, for example by exploiting data source integration), on the other hand they set the scene for a deeper study of individual characteristics and for segmentation analysis⁸.

The paper develops along this path of research by exploiting microdata derived from the statistical integration of the Italian labour force survey (LFS) sample with administrative records tracing formally regular jobs (hereafter summarised with ADMIN), where irregularity is flagged at individual level by comparing the employment status reported by independent sources. It is the first time that a micro integrated database is used to provide estimates of undeclared work for official statistics purposes. The integrated LFS sample (labelled with LFS-ADMIN⁹) 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 bias¹⁰. Investigation of heterogeneous nature of undeclared work, due both to individual strategies of firms and of workers and their interplay, as well as to the social contest in which they operate¹¹, has been furthermore made possible by the use of microdata. Those parts of the existing literature on undeclared work that treat the issue of its heterogeneous nature are firstly illustrated and discussed (par.1); an overview of LFS-ADMIN then follows with a description of undeclared work estimates (par. 2); some results obtained from modelling the probability of being in undeclared employment are then discussed (par. 3); then a segmentation of undeclared employment based on individual and job characteristics, as well as on the ADMIN traces of each individual, is provided (par. 4). Some conclusions are finally drawn.

1. Heterogeneity of undeclared work

Several domains of heterogeneity referred to undeclared work have been investigated in the existing literature and since a relatively long time. Portes *et al.* (1989) for instance recognize that if on the one hand informal economy has a universal character, since it is

⁷ They worked 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.

⁸ Eurobarometer (2007, 2013) conducted Europe-wide direct surveys on undeclared work commissioned by EU institutions. Nevertheless in European Commission (2014) it is also stated that such methods tend to under-report the extent of irregular work, particularly in specific domains. On this point see also, for example, Andrews *et al.* (2011).

⁹ The methodology adopted to build LFS-ADMIN has been developed by an ISTAT working group and it is described in AA.VV. (2015); for a concise description see par. 2. 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 starting with reference years 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). LFS-ADMIN estimates are now replied annually to update benchmark estimates: at the moment the delay is about t+17 months and it depends on the timing of ADMIN data. Some analyses on undeclared work derived from LFS-ADMIN are reported in ISTAT (2015, ch.4.1.2) and in De Gregorio *et al.* (2014).

¹¹ Under a theoretical approach these elements are developed in Pfau-Effinger (2009).

found in countries and regions characterized by very different economic systems and development achievements, on the other hand it is also quite a heterogeneous phenomenon, with large differences both between and within countries.

Heterogeneity can be examined under several perspectives and with varying degrees of complexity in the analytical framework. Simple descriptive statistics already clearly evidence the large variability in the incidence of irregular jobs across areas and economic activities. Labour demand and supply factors are clearly connected with the explanation of such heterogeneity: they involve the connections and vertical integration of irregular jobs with the formal side of the economy (“*comprising regulated economic units and protected workers*”)¹²; the characteristics of labour employed on the informal side, for instance concerning education and skill; the characteristics of individuals and of their environment, deriving from gender, citizenship, age, household structures and incomes, etc.; the general and local government attitudes towards the irregular sector, as summarised, for instance, by their effects on the functioning of local labour market and active labour market policies, on income distribution, on business structure, on the attitudes towards tax compliance and, more generally, on the quality of the social capital; by the legal and normative infrastructure that rules welfare and the functioning of the economy.

Further sources of heterogeneity derive from the intrinsic nature of undeclared work: the borders between regular and irregular labour input are not clearly distinguishable. As a matter of fact, undeclared labour input actually derives not only from straight irregular jobs (whether primary or secondary) but also from formally regular ones, due to unreported working time with partial evasion of social security and tax duties. The importance of this so called *grey* labour input might also be envisaged as the result of a partial adaptation to policies tailored to contrast purely hidden jobs¹³. It is interesting to notice that two main implications derive from this latter point. First, accurate estimates of grey and irregular labour input as a whole necessarily require actual working time information and estimates. Secondly, it becomes increasingly inadequate to represent irregularity as a headcount binary variable, while continuous or k-way categorical variables would better satisfy this purpose.

Put into a historical perspective, according to some authors the heterogeneity of irregular work has accompanied in the last decades the loss of homogeneity registered in the formal side of the economy, particularly in the industrial and services workforce and working conditions. The literature on dualism and industrial districts¹⁴ partly stresses these issues while analysing the substitution of solid vertical production processes with more fluid networks of horizontal activities. As a result, informal economy is intended as having reinforced such progress towards heterogeneity in working situations¹⁵.

The presence of informality in working conditions across different segment of working population is associated (if it is observed as a whole) with a wide range of employment opportunities and channels of access. Nevertheless these are usually segmented according

¹² Chen (2007).

¹³ 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) who deal with false part-time contracts in Italy. Boeri *et al.* (2002) focused instead on the fading borders between irregularity and unemployment.

¹⁴ See for instance the milestone provided by Berger *et al.* (1980).

¹⁵ This view is proposed in Capecchi (1989).

to the status of the worker: for instance citizenship, gender, age or else. It is widely recognized that women and ethnic minorities are more likely concentrated in low-paid, unskilled segments of irregular labour market. As a result, specific groups tend to be confined in specific markets.

Furthermore, heterogeneity is also seen as the result of individual choices based on the perceived advantages and disadvantages of not being protected by welfare coverages. In some cases the income differences between formal and informal employment may be not large, and in some segments workers could earn more money in informal than in formal employment¹⁶. Households conditions might also influence this balance, since access to social security might be assured by at least one member of the household.

Other dimensions of heterogeneity are deemed to be dependent on the degree of penetration of workers organizations (such as trade unions) and on elements tied to technological and sector-specific aspects. In some activities (for instance garment, construction or hotels and restaurants) informal labour practices is recognized to be used with the purpose of compressing labour costs. In the garment industry this attitude might reflect a deliberate (and possibly in some case successful) effort to increase competitiveness in the face of international competition; in other industries where international competition is irrelevant, as in the construction industry, undeclared work solutions come at the expense of organized labour. In hotels and restaurants organized labour has never been very strong: hence, immigrant labour and associated informal labour practices were not needed originally to lower firms' labour costs.

Firm size - as measured for instance in terms of number of employees - is also relevant in disentangling heterogeneity, since informality is observable in large and small firms, in capital-intensive and labour intensive industries. Heterogeneity in undeclared work goes with heterogeneity of their employers: it is increasingly apparent that “*by no means all businesses working on an off-the-books basis are wholly underground enterprises*”¹⁷. Most of those firms have been identified as firms officially belonging to the formal side of the economy but conducting a part of their trades on an *off-the-books* way¹⁸: their size ranges from micro businesses to small and medium sized enterprises.

Furthermore, the undeclared work related to self-employment confirms the above picture and multiplies the range of undeclared activities. In the literature, irregular self-employment has been envisaged as a sort of spontaneous answer to overregulated markets¹⁹. This neo-liberal approach has nevertheless been put into discussion more recently as the attention has been addressed towards the actual condition of those who are classified in self-employment. The works dedicated to the analysis of “dependent self-employment” has recognized that formal self-employment often hides economic dependence and this applies more frequently to undeclared businesses²⁰.

¹⁶ Williams *et al.* (2004).

¹⁷ Williams (2004a) p. 6.

¹⁸ Williams (2004b), Williams (2006).

¹⁹ There is a huge amount of literature on the issue of dependent self-employment, that is reviving in this last years due to the widespread need to provide official statistics with suitable definitions of employment in order to properly follow the increase in self-employment. See OECD (2000), Steinmetz *et al.* (1989), Linder *et al.* (1990), Muehlberger *et al.* (2007), Kautonen *et al.* (2010).

²⁰ See ILO (2002). Williams (2004b) distinguishes the “*micro-entrepreneurs starting-up fledging business ventures and using such work as a start-up strategy and on the other hand the more established self-employed who are serial users of underground work*”.

According to some authors, “*The emerging view is that the informal economy consists of various heterogeneous markets with different groups of individuals and firms engaged in a variety of informal activities, for diverse reasons and at varying pay/incomes*”²¹, and there is a wide scope for avoiding stereotypes and “*evaluating critically the representation of underground economy in advanced economies as comprised of marginalized populations working off the books as employees for wholly or partially underground businesses under exploitative conditions*”²². The idea behind this approach is that the adoption of a “thin” reading of irregular work only concerns a very particular segment of irregular labour market, while disregarding the multiplicity of forms and motivations assumed by irregular employment. Many studies in the last twenty years confirmed this approach by providing evidence that undeclared work is not necessarily concentrated in the weakest areas of the labour market or in the most deprived regions. Their finding is that a large part of irregular work derives from formally employed people, usually living in affluent regions. Without denying the weakest part of irregular employment and the presence of unscrupulous employers, it has been realized that irregular work does not necessarily imply low-paid jobs in an exploitative context. This depends mainly on the sector of activity, on the technology endowments and on the degree of organization of employers.

The conclusion that is usually drawn is that undeclared work derives from a combination of a plurality of factors²³: as a result individual causal factors alone do not provide useful explanations unless their interaction with local and environmental factors is appropriately taken into account. The causes and determinants of undeclared work are seen in particular as dependent on market relations (labour markets, goods and services markets and information markets), institutional relations (of citizens with public authorities and tax authorities), individual characteristics and other environmental factors²⁴. The causes within the above-mentioned categories all lead to various manifestations of undeclared work. This approach is helpful in policy design, in order to specifically tackle this mix of factors and circumstances: in other words policy should be tailor-made while there seems to be no scope for any standard recipe²⁵.

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

2.1. The integrated sample

Since the 1990s ISTAT has been producing estimates of undeclared work for national accounts purposes based on the integration of statistical sources. What became known in

²¹ Andrews *et al.* (2011, p.8).

²² This position is clearly stated in Williams *et al.* (2004, p. 2).

²³ Pfau-Effinger (2009).

²⁴ Arezzo (2013) reads undeclared work through the lenses of the theories of social capital.

²⁵ See Mateman *et al.* (2001), Renooy *et al.* (2004) and, for a detailed overview of this approach, Williams *et al.* (2004). It is worth noticing that these findings stem from the special attention that, as early as 1998, the European Commission decided to dedicate to the causes and consequences of undeclared work in the EU and to the possible policies to counter the phenomenon. For this purpose, the Commission issued the Communication on Undeclared Work and financed several research activities.

the literature as the “Italian approach”²⁶ was based on the statistical integration at domain level of estimates derived from business and administrative sources (covering regular jobs) and from household surveys and census data (covering regular and undeclared employment), within an analytical framework coherent with the objective of insuring exhaustiveness within the production boundary defined by the system of national accounts (SNA2003). The conceptual framework, widely accepted at international level, was based on the idea that undeclared work is the key to reconcile at domain level the estimates provided by independent sources.

More recently, the progress in the use of administrative sources, and in particular the availability of nearly exhaustive individual information on regular jobs²⁷, paved the way for further innovations in this area, with a passage from macro to microdata integration. The new approach is founded on the statistical integration of the LFS microdata with those of the administrative sources that trace regular employment²⁸. The general principle is to model for each individual in the integrated sample an estimate of her employment status and to flag it as undeclared where no validated administrative signals are available.

Italian LFS is a continuous survey with a yearly sample of more than 600 thousand interviews representative of individuals in the resident population²⁹: it provides monthly and quarterly figures for the main aggregates and yearly figures at NUTS3 level³⁰. The sampling design is rather complex, with two stages (municipalities are PSUs, households are FSUs), stratification of PSUs and rotation of FSUs³¹; within each NUTS3 domain, PSUs are selected with PPS sampling³². The sample is uniformly spread across all the weeks of the reference year: all territorial domains are represented in each month and in each of the four waves of the panel.

²⁶ “The ISTAT Analytical Framework relates the Non Observed Economy to the statistical problems to be addressed by national accountants so as to identify the origins of the lack of exhaustiveness and their impact on the statistical system”, OECD (2002, pag.42). More details are provided by Calzaroni (2000).

²⁷ The development of a linked employed-employee database by ISTAT has been of the utmost importance to spur the adoption of micro level statistical integration of survey and administrative sources.

²⁸ It has been developed and introduced to support the estimates of undeclared work at individual level in occasion of 2011 national accounts benchmark. Full details of the new approach are reported in AA.VV. (2015).

²⁹ Italian LFS survey is run within the legal framework set by the corresponding EU statistical regulations. The principal legal act is the Council Regulation (EC) No 577/98 (see the most recent quality report in Eurostat 2015). It should be reminded that 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 and cover the labour input of non-residents.

³⁰ NUTS3 level of the territorial units classification (last version is 2014) corresponds to the more than one hundred “province” in which Italy is actually split. This level is 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.

³¹ Here PSUs and FSUs are the usual acronyms for primary and final sampling units. Households follow a 2-2-2 rotation scheme: they participate to the survey for two consecutive quarters, then they temporally exit from the sample for the following two quarters, and then come back in the sample for two quarters. This produces a 50% overlap of the sample between a quarter and the previous one and a 50% overlap between a quarter and the same quarter of the previous year. For details see Discenza *et al.* (2014) and Di Consiglio *et al.* (2014).

³² PSUs are stratified according to the demographic size. Large municipalities are always included in the sample; the others are selected within each stratum with probability proportional to the size of the resident population.

The individual tax code is the key adopted for record linkage with ADMIN microdata³³. The LFS sample weights are here used for modelling statistical integration, for running descriptive analyses on the integrated sample (further in this section), and in the analyses of unit level data (sections 3 and 4). The choice of a survey-weighted estimator in the logistic model considered in section 3 derives mainly from the complexity of the LFS sample design which make it most likely not ignorable. It is well known that maximum likelihood estimators become biased and inconsistent when the sampling design is informative³⁴; nevertheless, the use of weighted estimators can result in substantial loss of efficiency which is expected to be larger the smaller is the sample size and the larger the variation of the sampling weights³⁵. Actually, the LFS sample is quite large and the variability of sampling weights appears relatively restrained (cv=75%): it is largely determined by PSUs inclusion probabilities, while there is no evidence that the response variable – i.e. the nature of employment (regular vs. undeclared) - has noticeable influence on such probabilities. Despite the complexity of the LFS sample, several factors explain the strong equivalence which has been found between the results obtained from weighted and unweighted estimators: this evidence seems due to the structure and size of the LFS sample that contribute to smooth the effects due to the likely informative nature of the design.

LFS-ADMIN integration consisted in the estimate of the actual employment status for each record in the sample based on a statistical model aimed at reconciling the information independently gathered by LFS and ADMIN. Potential sources of incoherence may in fact derive from time-related and definitions-related issues. As concerns the former, it is well known that the employment status recorded by LFS for each individual is in fact referred to a particular week in the year (the “reference week”). On their part ADMIN data are a set made of several distinct sources³⁶ characterised by varying degrees of precision in detailing the characteristics – and namely the time profile – of actual labour input. Differences concern in particular their accuracy in detailing the dates of the employment status and in focussing the events taking place in LFS reference week: to simplify, ADMIN sources are usually very accurate for employees but less precise for self-employment.

As for definitions, the employment status recorded 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

³³ This deterministic procedure of record linkage has been regarded as highly reliable and – given the time constraints – a lower priority has been given to the evaluation of linkage errors: of course, such aspects needs further attention to take properly into account the fact that for less than 5% of the individuals in LFS sample the tax code could not be validated through ISTAT Population register. In these cases the tax codes have been imputed based on the structural characteristics of the individual (*inter alia*, gender, age, territory, citizenship, LFS employment status). This choice has been driven by the need to consider the LFS as a whole, in order to adopt the original sample weights and to compare directly original LFS estimates with those of the integrated sample. The individuals with imputed tax code were not considered in the estimation of the parameters of the integration model described below. See AA.VV. (2015) for more details.

³⁴ See Pfeffermann (1993) and Skinner in Chambers *et al.* (2003, ch.6).

³⁵ Pfeffermann (1993), Chambers *et al.* (2003, ch. 6). See also Kish (1990).

³⁶ ADMIN data derive mainly from social security sources on employees of private enterprises in industry and services (INPS-EMENS), in recreation (ENPALS), agriculture (INPS-DMAG), of domestic personnel (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 agriculture (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).

rules that do not necessarily match ILO standards: for instance, it only refers to labour input with formal traces and thus excludes by definition entirely undeclared jobs; furthermore it may include false positives.

Tackling reconciliation thus implies the adoption of methods to detect, measure and correct the biases affecting both sources: notably, the possible under-coverage of employment and particularly of secondary jobs by LFS³⁷, ADMIN over-coverage of regular jobs and ADMIN lack of coverage of irregular work³⁸. By adopting source-dependent error models, the actual strength of the employment signals conveyed by each ADMIN source is evaluated through the comparison with LFS employment status³⁹. This process is aimed at filtering ADMIN sources in order to render them homogeneous providers of signals concerning the actual employment status of the individual in the reference week. Given the validated ADMIN employment status, the probability of LFS employment under-coverage is predicted at individual level. Irregular jobs have finally been defined as employment spells unmatched in the reference week with validated ADMIN signals. The integrated dataset lists the jobs performed by the individuals in the LFS sample who are actually employed, with further details concerning the order of the job (whether primary, secondary, etc.), the regularity status, the economic activity (4-digit NACE), the number of weekly actually worked hours, the type of employment, the tasks and duties undertaken in the job⁴⁰, the business register data on the employer and the rest of LFS information collected through the survey questionnaire⁴¹. Integrated job data are thus combined with the personal characteristics of the worker and with the whole profile of his yearly ADMIN records⁴². Since the focus here is on employment, only the individuals in employment according to the integrated estimates⁴³ have been selected from LFS-ADMIN, by considering only their

³⁷ 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. (2015) for a deeper insight of this issue.

³⁸ ADMIN over-coverage is source dependent: the accuracy of the dating of actual labour input is the core issue. As a matter of fact, a lack of precision mainly affects the sources on self-employment. On the contrary, those on employees are usually very precise and report duration and dates of labour contracts. The probabilistic approach adopted for LFS-ADMIN integration is fully described in AA.VV. (2015) and it is inspired by the recent ESS literature on data integration. See also De Gregorio, Filippini *et al.* (2014). Previous research by the ESSnet on data integration drove this approach: see also García Martínez (2011), Hochfellner (2011), Kuijvenhoven *et al.* (2011), Linder *et al.* (2012), Zhang (2012). Pavlopoulos *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 also be explored to measure irregularity.

³⁹ It is assumed that P_k^R is the probability for the individual k in the LFS sample to have a “true” regular employment status ($y_k^R=1$) conditional to the values assumed by the auxiliary variables in ADMIN (A_k) and in LFS (S_k) and given the ADMIN employment status ($a_k=1$):

$$P_k^R = \Pr(y_k^R = 1 | a_k = 1, A_k, S_k)$$

Based on an estimate of P_k^R and on the LFS error model, individual predictions of irregular employment status undeclared to LFS are derived. See AAVV(2015) and De Gregorio, Filippini *et al.* (2014).

⁴⁰ They are coded through ISCO nomenclature.

⁴¹ In the case of irregular jobs, the information is derived mostly from the answers to the LFS questionnaire and eventually 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. (2015) for further details.

⁴² In perspective, ADMIN data can be organized longitudinally and individual regular histories can be used more efficiently to outline and detect irregularity.

⁴³ They include thus all the individuals in LFS sample who are in employment according to LFS plus the remaining individuals rescued throughout ADMIN signals and LFS under-coverage estimates.

primary job (be it regular or irregular)⁴⁴: considering years 2010 and 2011 together, the total sub-sample consists of about 480.000 individuals, 48.000 of which with an irregular primary job⁴⁵.

2.2. Some descriptive evidence

Within the definition given in advance, in this framework undeclared work here refers only to “fully undeclared work” of resident population: both “partially undeclared work” (also called “under-declared work” as, for instance, a full-time employee officially registered as a part-time one) and fully undeclared work of non-resident population are not considered here.

The descriptive statistics reported in Table 1 derive from the use of the LFS-ADMIN integrated sample with the original LFS sampling weights. Given the assumptions concerning the deterministic nature of record linkage, ADMIN data and the results of the integration process can be regarded *de facto* as mere extensions of the LFS questionnaire.

LFS-ADMIN estimates for the whole period 2010-2011 confirm some expected characteristics of irregular employment already highlighted by other independent estimates⁴⁶. The incidence of undeclared 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; the households structure and the role of the individual within the household both play a non-secondary role; the presence of other irregular workers in the household is also associated with larger irregularity rates.

⁴⁴ According to the ESA regulation, 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 job.

⁴⁵ These individuals originate about 55 thousand secondary jobs, 8.000 of which correspond to irregular jobs. All these figures are very stable between 2010 and 2011.

⁴⁶ 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. Indicators of undeclared work by segment. Two-year 2010-2011 (%)

Segment	Rate (a)	Share on total(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: Agriculture	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) Undeclared employment as percentage of total employment in the segment.

(b) Undeclared employment in the segment as percentage of total undeclared employment.

From statistical integration do emerge further characteristics of undeclared workers: for instance, it is important to notice that they are frequently traced in ADMIN during the reference year though not in the reference week. 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 imply some switching from regularity, suggesting again that the treatment of undeclared

work should be followed through continuous or at least multi-modal variables⁴⁷: this point is also supported by the circumstance that an important share of undeclared workers reports to the LFS to be working in large local units⁴⁸. These aspects open the way for specific analyses dedicated to undeclared outsourcing of services by larger enterprises. Furthermore, irregular jobs are also 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, the integrated sample confirms also that heterogeneity lays behind them⁴⁹. 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.

3. Modelling the probability of being an undeclared worker

The probability for a person in employment to work undeclared can be analysed from different perspectives. In this section, the individual characteristics (such as age, gender, household, citizenship, etc.) have been primarily used as independent variables: other variables have been progressively introduced to summarize local context effects.

A battery of NUTS3 (i.e. province) level labour market indicators is used to monitor the effects induced by actual local market and active policy conditions on individual profiles: the activity and unemployment rates, the contact rate with public or private job centers⁵⁰, the coexistence rates of the so called *grey area* with official labour force and of potential employment⁵¹ and with total employment⁵². For Italian LFS the NUTS3 coefficients of variation for the unemployment rate ranged from 3.7 to 30.8 in 2011⁵³, with a median of 10.2% and 5th and 95th percentiles respectively equal to 5.1 and 18.8%⁵⁴.

A proxy of the local attitudes towards tax compliance was adopted, based on the

⁴⁷ Such evidence might encourage the adoption of a fuzzy variable approach to target irregularity. This approach has been for instance already 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).

⁴⁸ The afore mentioned "Commissione Giovannini" strongly emphasised the need to provide sound estimates of the so called *grey area* of undeclared work.

⁴⁹ This point is also stressed in the contribution of Cappariello *et al.* (2009).

⁵⁰ 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 the 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.

⁵¹ As defined by the sum of unemployed and *grey area*.

⁵² 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.

⁵³ Eurostat (2015), ch.9.

⁵⁴ 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.

partition of the NUTS3 levels into eight clusters provided in the DBGEO database developed by the Italian tax authority⁵⁵. 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) – has been used to introduce the demand side of the irregular labour market.

By using the nature of employment - whether regular or undeclared - as the response variable, a logistic model (1) has been run to estimate the probability of undeclared work in function of the above mentioned sets of variables. Several specifications have been tried, changing the sets of variables, the interactions and the model groups⁵⁶. What follows is the general simple effect version:

$$\text{logit}(IRR_{i(gk)}) = \alpha + P'_{i(gk)}\beta + X'_{(gk)}\gamma + W'_{i(gk)}\lambda \quad (1)$$

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.

Employment is generally characterized by gender differences, for example, in participation, skills, earnings, types of work and working conditions⁵⁷. Gender is here considered as an element of heterogeneity of undeclared work as well⁵⁸: so, as an alternative to including gender into the model as a dummy variable including its interactions with other predictors, it has been preferred the adoption of two distinct models, one for females and one for males, in order to investigate different aspects of the gender dimension of undeclared work.

Three simple effects models were run separately on the two genders, and some results are reported hereafter: model A uses only P variables, model B introduces the X set and model C adds 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%⁵⁹. 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

⁵⁵ The clustering is based on variables concerning tax behavior, criminality, consumption patterns, business structure, technological development, transport infrastructure, characteristics of taxpayers (see for more details “*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 aprile 2014.). A first cluster, labelled *All right*, joins a high life standard with appreciable tax compliance; *the Equilibrist*, groups small NUTS3 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 NUTS3 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.

⁵⁶ The main results obtained throughout alternative specifications do not differ substantially. Models with weighted and unweighted observations have also been tested, without appreciable differences. All the data reported in this work derive from the use of weighted observations.

⁵⁷ World bank (2012).

⁵⁸ Renooy *et al.*(2004), Capecchi (1989).

⁵⁹ See the tables A.1-A.3 in the Appendix for details on model fit and estimates.

irregular with an odds ratio in model A larger than 1.4 points as 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 males⁶⁰ - 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 elder 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 genders do emerge if the effect of household structure is accounted for. Men living alone have a relatively higher probability of undeclared work, with an odds ratio that doubles that of adults living with a partner and a son (the benchmark less "at risk of irregularity"). The corresponding odds ratio for women is far lower and this might be due to the conditions laying behind the choice of living alone⁶¹. 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 earner operates quite differently according to whether this additional income 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 level of education (such as the completion of secondary schools)⁶².

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 labour market policies when associated to high unemployment rates may partially translate into a higher participation in the irregular side of the market. This does not seem to be the case for men, for whom higher contact rates with job centers reduce the risk of irregularity. Tax compliance attitudes cope well with explaining undeclared work for both genders, in particular when territories are included in the *Total risk* cluster. The effect associated with the structure of regular business deserves some attention. For males, the higher the relative weight (in terms of regular employees) of "risky" sectors such as agriculture and construction, the lower the probability of being employed in undeclared jobs: this might be connected with the emersion of previously undeclared activities, as a likely reaction to policies aimed at fighting this phenomenon⁶³. On the contrary, the relative weight of regular employees in households services and in microenterprises seems related to higher irregularity risks.

The introduction into the analysis of details on the characteristics of the 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 men. 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.

⁶⁰ See the table A.4 for significance tests on the differences in estimates of logistic model coefficients by gender.

⁶¹ It should be noticed that the household here described derives from administrative population registers, and might not coincide with the actual "economic household".

⁶² 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 a sign of the weakness of local markets and policies to meet this segment of labour supply.

⁶³ This effect does not seem anyway to be at work 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 undeclared. Should his ISCED level 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 at 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 his 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.

An appropriate evaluation of these results needs a consideration of the error associated to the integration model⁶⁴. 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.

⁶⁴ De Gregorio, Filippini *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), Pavlopoulos *et al.* (2012), Zhang (2012). Replication techniques and bootstrapping have been used by De Gregorio, Filippini *et al.* (2014) in order to validate these estimates of irregular labour input for national accounts purposes: they provided encouraging results (see also as a references Wolter 2007, Kuijvenhoven *et al.* 2011).

Table 2. Predicted probability of being in undeclared employment, by gender, profile and model (%)

Profile	Age	Role and household 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		
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-Apr	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

4. A segmentation of undeclared 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 clustering⁶⁵: the focus now is more strictly on the irregular job and on the sector specific features related to the use of undeclared 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 signals⁶⁶.

With nearly 50 variables and 150 modalities the first ten eigenvalues accounted for about 40% of total inertia⁶⁷. 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 activity. The results from the overall sample are reported hereafter. In general, the first component (6.5% of total inertia) opposes two poles that could be summarized as “*unskilled blue collars*” vs. “*skilled self-employed*”⁶⁸. 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*”⁶⁹. 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:

⁶⁵ Fuzzy clustering could eventually be explored in order to take account of the above mentioned fuzziness of the concept of irregularity.

⁶⁶ The net monthly income declared to LFS, the hours actually and normally worked, the number of secondary jobs have been used as illustrative variables.

⁶⁷ 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.

⁶⁸ 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.

⁶⁹ 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

they are relatively old and mostly Italians and from the South. Self-employment in trade activities is also well represented in this cluster and in cluster 1 (alike business services), where education level is higher and where central and northern regions and male employment have a relatively higher presence.

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 the 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 side the patterns of irregularity look somehow ADMIN-dependent in the sense that they seem to have been moulded by sector specific habits and needs and by local influences.

Table 3. Clusters of individuals in undeclared employment, by NACE, gender, age and citizenship (% distrib. and specialization rates by segment)

Cluster	Short description	Distr.	Specialization rates											
			NACE											
			Total	Agriculture	Industry	Construction	Trade Horeca	Business services	Household services	Women	Young 15-34	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	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 edu., 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 occ., Low edu., South, Parents, Weak ADMIN	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	Female, 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, Male, Med.-young, Low edu.	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

Concluding remarks

The availability of household survey microdata is essential for disentangling the complexity of undeclared employment. The main challenge is represented by the micro-level indirect detection of irregular job holders and by 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.

This work tests the possibilities offered by the integrated LFS-ADMIN sample developed by ISTAT to support national accounts benchmark estimates. The descriptive analyses seem to confirm the results derived from previous research adding more details on irregular employment, especially with reference to the heterogeneous characteristics of individuals and of their environment. A measurement of the effect of those characteristics on the probability of being in undeclared employment 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 undeclared according to whether the local labour markets are or are not endowed with appreciable inclusion capabilities: huge inactivity rates, 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 undeclared work. This aspect and the causal links need further investigation based on tailored methodological approaches.

The segmentation of undeclared 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 approaches to contrast irregularity, based on active policies and where local conditions should receive greater attention. Although these suggestions need further assessment, they seem to confirm the results of previous researches on this subject, especially those conducted more recently and spurred by the European Commission. In particular, Italian results clearly identify the heterogeneous nature of undeclared work and help to reject any stereotyped view as a fully marginalised segment of the labour market. Nevertheless, they confirm also that such heterogeneity derives mainly from the interaction of labour demand with local labour market performances, while the weakness of the conditions of irregular labour supply (education, age, gender, skills, household conditions) accompany the large part of the individuals on this side of the market. For this reason, these results also evidence that there is a large scope for policies, in order to recover locally the human and social capital lost in undeclared activities.

Though encouraging, the approach based on statistical integration of independent sources also deserves some further deepening under several profiles: they mainly concern definitions and methodological issues. The boundaries of irregularity need in particular to be accurately scrutinised. In the developments shown here, undeclared employment corresponds to work that is 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 hypothesis that the LFS records only legal businesses, although possibly undeclared: 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 actual working time both from households and business statistics sides⁷⁰.

The methodological aspects are those who appear more promising, both for the statistical integration process 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 might be worth to be tested. The use of more sophisticated approaches based on logistic regression could also 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 to those concerning income and tax registers: this perspective appears extremely appealing in order to provide income and labour input estimates in a same methodological environment. 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 integration, 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 make a deeper use of the information in the whole set of ADMIN data, which cover the universe of the formally regular jobs the present population is engaged in.

⁷⁰ Baldi *et al.* (2013).

References

- AA.VV. 2015. Soluzioni metodologiche per l'utilizzo integrato delle fonti statistiche per le stime dell'occupazione. *ISTAT working paper*, n.19. Roma: ISTAT
- Andrews D., Caldera Sánchez A., Johansson Å. 2011. Towards a better understanding of the informal economy. Economics department working papers no.873. Paris: OECD
- Arezzo M.F. 2013. Social capital and undeclared work: an empirical analysis in Italy from 1998 to 2008. Working Paper del Dipartimento di Metodi e Modelli per l'Economia il Territorio e la Finanza, n. 112. Roma: Università La Sapienza
- Baccini A., L. Castellucci e M. Vasta. 2003. Lavoro sommerso: il caso della Toscana. *Studi e note di economia*, n. 1, Firenze
- Baldassarini A. 2001. Non regular foreign input of labour in the new national accounts estimates. OECD meeting of national accounts experts. Paris, 9-12 October. Paris: OECD
- Baldi C., De Gregorio C., Giordano A., Pacini S., Solari F., Sorrentino M. 2013. 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
- Berger S., Piore M.J. 1980. *Dualism and Discontinuity in Industrial Societies*. Cambridge: Cambridge University press
- Betti G., Ferretti C., Gagliardi F., Lemmi A., Verma V. 2009. Proposal for new multidimensional and fuzzy measures of poverty and inequality at national and regional level. Working Paper del Dipartimento di Metodi Quantitativi, n. 83, November. Siena: Università di Siena
- Boeri T., Garibaldi P. 2002. Shadow activity and unemployment in a depressed labour market. *CEPR discussion paper*, n.3433. Roma: CEPR
- Calzaroni M. 2000. 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
- Capecchi V. 1989. The Informal Economy and the Development of Flexible Specialization in Emilia-Romagna. In: Portes *et al.* *The informal economy. Studies in advanced and less developed countries*. London: The Johns Hopkins University Press
- Cappariello R., Zizza R. 2009. Dropping the books and working off the books. *Temi di discussione*, n. 702. Roma: Banca d'Italia
- Cascioli R. 2006. 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
- Chambers R. L., Skinner C. J. (eds.). 2003. *Analysis of survey data*. John Wiley & Sons, Ltd.
- Chen M.A. 2007. Rethinking the Informal Economy: Linkages with the Formal Economy and the Formal Regulatory Environment. *DESA Working Paper* No. 46. New York: United Nations

- Colubi A., Coppi R., D'Urso P., Gil M.A. 2007. Statistics with fuzzy random variables. *Metron International Journal of Statistics*, vol. LXV, n. 3, pp. 277-303. Roma
- De Gregorio C., Filippini D., Martini A., Rocchetti I. 2014. A comparison of sample and register based survey: the case of labour market data. Paper presented at Q2014 Conference. Vienna: Statistik Austria
- De Gregorio C., Giordano A. 2014. "Nero a metà": contratti part-time e posizioni full-time fra i dipendenti delle imprese italiane. *ISTAT working paper*, n.3. Roma: ISTAT
- Di Consiglio L., Loriga S., Martini A., Ranaldi R. (et al.). 2014. The Italian LFS sampling design: recent and future developments. 9th Workshop on Labour Force Survey Methodology, Rome 15-16 May
- Disenza A., Loriga S., Martini A. 2014. The Italian Labour Force Survey consistency framework. 9th Workshop on Labour Force Survey Methodology, Rome 15-16 May
- Eurobarometer. 2007. Undeclared Work in the European Union. Final report. Special Eurobarometer 284, Wave 67.3. TNS Opinion & Social
- Eurobarometer. 2013. Undeclared Work in the European Union. Final report. Special Eurobarometer 402, Wave EB79.2. TNS Opinion & Social
- European Commission. 2007. *Stepping up the fight against undeclared work*. Communication from the Commission to the Council, the European parliament, the European economic and social committee and the Committee of the regions (COM(2007) 628 final). Bruxelles
- European Commission. 2014. *Employment and Social Developments in Europe 2013*. Directorate-General for Employment, Social Affairs and Inclusion Directorate A. Bruxelles: European Commission
- Eurostat. 2015. *Quality report of the European Union Labour Force Survey 2014*. Luxembourg: Eurostat
- García Martínez M.A., Galindo J.O. 2011. 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
- GHK, Fondazione G. Brodolini. 2009. *Study on indirect measurement methods for undeclared work in the EU. Final Report*. European Commission, Directorate-General Employment, Social Affairs and Equal Opportunities.
- Gobbi G., Zizza R. 2007. Does the underground economy hold back financial deepening? Evidence from the Italian credit market. *Temi di discussione*, n. 646. Roma: Banca d'Italia
- Hochfellner D. 2011. Dealing with inconsistencies in the integration of multiple register based data sources. *NTTS paper*. Luxembourg: Eurostat
- ILO. 2002. *Decent work and the informal economy*. Geneva: International Labour Organization
- ILO. 2010. *Labour inspection in Europe: undeclared work, migration, trafficking*. Geneva: International Labour Organization
- ILO. 2013. *The Informal Economy and Decent Work A Policy Resource Guide supporting transitions to formality*. Geneva: International Labour Organization

- ISFOL. 2007a. 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
- ISFOL. 2007b. Le politiche per l'emersione del lavoro nero e irregolare. Quadro istituzionale e normativo. *Monografico dell' Osservatorio Istituzionale e Normativo*, n. 11. Roma: ISFOL
- ISFOL. 2011. Dimensioni e caratteristiche del lavoro sommerso/irregolare in agricoltura. *Area "Sistemi Locali e Integrazione delle Politiche"*. Roma: ISFOL
- ISTAT. 2015. *Rapporto annuale 2015 – La situazione del Paese*. Roma: ISTAT
- Kish L. 1990. Weighting: why, when and how? *Proceedings of the section on survey research methods*. American Statistical Association, 121-130
- Linder M., Houghton J. 1990. Self-employment and the petty Bourgeoisie: comment on Steinmetz and Wright. *American Journal of Sociology*, vol. 96 n. 3 (Nov.), pp. 727-735
- Lucifora C. 2003. *Economia sommersa e lavoro nero*, Bologna, Il Mulino, pagg.154
- Kautonen T., Down S., Welter F., Vainio P., Palmroos J., Althoff K., Kolb S. 2010. "Involuntary self-employment as a public policy issue: a cross-country European review. *International journal of entrepreneurial behavior & research*, vol.16 n.2.
- Kuijvenhoven L., Scholtus S. 2011. Bootstrapping Combined Estimators based on Register and Sample Survey Data. *Statistics Netherland discussion paper*, n.23. Den Haag: CBS
- Linder F., van Roon D., Bakker B. 2011. 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
- Mateman S., Renooy P.H. 2001. Undeclared labour in Europe: towards an integrated approach of combating undeclared labour. Amsterdam: Regioplan Research Advice and Information
- MEF. 2011. *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
- Muehlberger U. 2007. *Dependent self-employment. Workers on the border between employment and self-employment*. Palgrave Macmillan: New York, p.217.
- OECD. 2000. The partial renaissance of self-employment. In: *OECD employment outlook*, ch.5. OECD: Paris, pp. 155-199.
- OECD. 2002. *Measuring the non-Observed Economy. A Handbook*. Paris: OECD
- Pavlopoulos D., Vermunt J.K. 2012. Measuring temporary employment. Do survey or register data tell the truth? EALE 2012 conference. Bonn
- Pfau-Effinger B. 2009. Varieties of Undeclared Work in European Societies. *British Journal of Industrial relations*, 47:1 (March)
- Pfeffermann D. 1993. The Role of Sampling Weights When Modeling Survey Data. *International Statistical Review*. Aug., pp. 317-337
- Portes A., Castells M., Benton L.A. 1989. *The informal economy. Studies in advanced and less developed countries*. London: The Johns Hopkins University Press

- Renooy P., S. Ivarsson O. van der Wusten-Gritsai, R. Meijer. 2004. *Undeclared Work in an Enlarged Union: an analysis of shadow work and an in-depth study of specific items*. Brussels: European Commission
- Steinmetz G., Wright E.O. 1989. The fall and rise of the petty bourgeoisie: changing patterns of self-employment in the postwar United States. *American Journal of Sociology*, vol. 94 n. 5 (Mar.), pp. 973-1018
- Williams C.C. 2004a: Geographical Variations in the Nature of Undeclared Work. *Geografiska Annaler*, 86 B (3): 187–200
- Williams C.C. 2004b. *Small Businesses in the Informal economy: making the transition to the formal economy*. London: Small Business Service
- Williams C.C. 2006. *The Hidden Enterprise Culture*. Cheltenham: Edward Elgar
- Williams C.C. 2007. Tackling Undeclared Work in Europe: Lessons from a Study of Ukraine. *European Journal of Industrial Relations*, SAGE Publications, 2007, 13 (2), pp.219-236
- Williams C.C. 2010. 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
- Williams C.C., Windebank J. 2004. The heterogeneity of the underground economy. *International Journal of Economic Development*, Vol. 6, n. 2, pp. 1-23
- Wolter K.M. 2007. *Introduction to variance estimation*. Springer (2nd edition)
- World Bank. 2012. *Gender Equality and Development*. World development report 2012. Washington DC
- Zhang L.-C. 2012. Topics of statistical theory for register-based statistics and data integration. *Statistica Neerlandica*, vol. 66, nr. 1, pp.41-63. Den Haag: CBS
- Zizza R. 2002. Metodologie di stima dell’economia sommersa: un’applicazione al caso italiano. *Temi di discussione*, n. 463. Roma: Banca d’Italia

Appendix

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

Model Fit Statistics	MALE				FEMALE			
	No Intercept	Model A	Model B	Model C	No Intercept	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

Effect	D F	MALE			FEMALE		
		Mode IA	Mode IB	Mode IC	Mode IA	Mode IB	Mode IC
CITIZENSHIP	2	<.00 01	<.00 01	<.00 01	<.00 01	<.00 01	<.00 01
AGE CLASS	5	<.00 01	<.00 01	<.00 01	<.00 01	<.00 01	<.00 01
HOUSEHOLD STRUCTURE AND ROLE	1	<.00 01	<.00 01	<.00 01	<.00 01	<.00 01	<.00 01
EDUCATION	4	<.00 01	<.00 01	<.00 01	<.00 01	<.00 01	<.00 01
OTHER HOUSEHOLD REGULAR INCOMES	1	<.00 01	0.060	0.016	01	0.000	01
OTHER HOUSEHOLD IRREGULAR INCOMES	1	<.00 01	<.00 01	<.00 01	<.00 01	<.00 01	<.00 01
LABOUR MARKET (by Gender) Fact.1 (a)	1		<.00 01	<.00 01	<.00 01	<.00 01	<.00 01
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		<.00 01	<.00 01	<.00 01	<.00 01	<.00 01
REGULAR EMPLOYMENT STRUCTURE (by Gender) Agriculture	1		0.613	01	<.00 01	<.00 01	<.00 01
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	<.00 01	<.00 01	<.00 01
REGULAR EMPLOYMENT STRUCTURE (by Gender) Business services	1		0.061	0.204	01	01	
REGULAR EMPLOYMENT STRUCTURE (by Gender) Household services	1		<.00 01	<.00 01	<.00 01	<.00 01	<.00 01
SHARE OF EMPLOYMENT IN MICROENTERPRISES	1		0.000	0.001	0.525	0.891	
TYPE OF JOB (Employee/Self-employed)	1			0.000		<.00 01	<.00 01
IRREGULAR JOB NACE CODE	5			<.00 01		<.00 01	<.00 01

(a) Unemployment, inactivity and grey area vs. virtuous labour market.

(b) Unemployment and placement vs. grey area.

(c) Placement.

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

Variables	Modalities	MALE						FEMALE					
		Model A		Model B		Model C		Model A		Model B		Model C	
		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 (ref=Extra EU)	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) ***	
	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 (ref=75+)	15-24	-0.088 (0.0267) *		-0.119 (0.0271) ***		-0.057 (0.0278) *		0.098 (0.0333) **		0.103 (0.0338) **		0.327 (0.0345) ***	
	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) ***	
	35-54	-0.716 (0.0116) ***		-0.748 (0.0162) ***		-0.703 (0.0165) ***		-0.748 (0.0232) ***		-0.797 (0.0235) ***		-0.699 (0.024) ***	
	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) ***	
	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 (ref=Other)	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.	
	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) ***	
	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) ***	
	Parent(1Parent&1son)	-0.197 (0.0822) *		-0.066 (0.0853) n.s.		-0.085 (0.0837) n.s.		-0.191 (0.0369) ***		-0.050 (0.0375) n.s.		-0.026 (0.0378) n.s.	
	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.	
	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) ***	
	Spouse(2with other comp.)	-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.	
	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) ***	
	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) ***	
	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.	
	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 (REF=ISCED=6)	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) ***	
	ISCED 2	0.115 (0.0124) ***		0.099 (0.0126) ***		0.109 (0.0129) ***		0.144 (0.0137) ***		0.146 (0.0138) ***		0.158 (0.0142) ***	
	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) ***	
	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 HH INCOMES (ref=None)	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) ***	
	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) ***	

Table A.3 (continued). Parameter estimates, by model and model group

Variables	MALE						FEMALE																																																																																																																																																																																																																																																																																																																
	Model A		Model B		Model C		Model A		Model B		Model C																																																																																																																																																																																																																																																																																																												
	Est.	St.Err.	Pr.	Est.	St.Err.	Pr.	Est.	St.Err.	Pr.	Est.	St.Err.	Pr.																																																																																																																																																																																																																																																																																																											
LABOUR MKT (by gender)													Factor1(a)	0.093	(0.00729)	***	0.093	(0.00733)	***	0.076	(0.00771)	***	0.073	(0.00778)	***	Factor2(b)	-0.006	(0.0092)	n.s.	-0.006	(0.00923)	n.s.	0.025	(0.0122)	*	0.028	(0.0123)	*	Factor3(c)	-0.055	(0.0147)	**	-0.058	(0.0148)	**	0.000	(0.0179)	n.s.	0.007	(0.018)	n.s.	Equilibrat	0.007	(0.024)	n.s.	-0.001	(0.0241)	n.s.	-0.025	(0.0265)	n.s.	-0.046	(0.0267)	n.s.	Industrial	-0.183	(0.0277)	***	-0.185	(0.0278)	***	-0.149	(0.0286)	***	-0.142	(0.0289)	***	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 (ref=All right)	0.055	(0.0309)	n.s.	0.054	(0.0311)	n.s.	0.105	(0.0306)	**	0.071	(0.0309)	*	Not angels	0.031	(0.0301)	n.s.	0.035	(0.0303)	n.s.	0.097	(0.0344)	**	0.113	(0.0347)	**	Risky habits	-0.053	(0.0324)	n.s.	-0.054	(0.0325)	n.s.	-0.100	(0.0375)	**	-0.119	(0.0378)	**	Total risk	0.274	(0.0335)	***	0.281	(0.0337)	***	0.243	(0.0378)	***	0.269	(0.0381)	***	Agriculture	-0.001	(0.0026)	n.s.	-0.011	(0.00263)	***	0.033	(0.00359)	***	0.023	(0.00366)	***	Construction	-0.011	(0.00394)	**	-0.014	(0.00396)	**	-0.017	(0.0135)	n.s.	-0.021	(0.0137)	n.s.	Trade&Horeca	0.002	(0.00342)	n.s.	0.001	(0.00345)	n.s.	0.018	(0.00309)	***	0.018	(0.00313)	***	Businessservices	0.005	(0.00266)	n.s.	0.003	(0.00268)	n.s.	0.019	(0.00317)	***	0.018	(0.00321)	***	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 (ref=Employees)													Self-employed	-0.030	(0.00788)	**	-0.030	(0.00788)	**	-0.036	(0.00884)	***	-0.356	(0.00884)	***	Agriculture				0.487	(0.0196)	***				0.515	(0.0263)	***	Industry				-0.396	(0.0162)	***				-0.137	(0.0243)	***	IRREGULAR JOB NACE (ref=HH services)				0.130	(0.0156)	***				-0.088	(0.0547)	n.s.	Trade&Horeca				-0.216	(0.0147)	***				-0.213	(0.0187)	***	Business services				-0.304	(0.0158)	***				-0.344	(0.0213)	***
Factor1(a)	0.093	(0.00729)	***	0.093	(0.00733)	***	0.076	(0.00771)	***	0.073	(0.00778)	***																																																																																																																																																																																																																																																																																																											
Factor2(b)	-0.006	(0.0092)	n.s.	-0.006	(0.00923)	n.s.	0.025	(0.0122)	*	0.028	(0.0123)	*																																																																																																																																																																																																																																																																																																											
Factor3(c)	-0.055	(0.0147)	**	-0.058	(0.0148)	**	0.000	(0.0179)	n.s.	0.007	(0.018)	n.s.																																																																																																																																																																																																																																																																																																											
Equilibrat	0.007	(0.024)	n.s.	-0.001	(0.0241)	n.s.	-0.025	(0.0265)	n.s.	-0.046	(0.0267)	n.s.																																																																																																																																																																																																																																																																																																											
Industrial	-0.183	(0.0277)	***	-0.185	(0.0278)	***	-0.149	(0.0286)	***	-0.142	(0.0289)	***																																																																																																																																																																																																																																																																																																											
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 (ref=All right)	0.055	(0.0309)	n.s.	0.054	(0.0311)	n.s.	0.105	(0.0306)	**	0.071	(0.0309)	*																																																																																																																																																																																																																																																																																																											
Not angels	0.031	(0.0301)	n.s.	0.035	(0.0303)	n.s.	0.097	(0.0344)	**	0.113	(0.0347)	**																																																																																																																																																																																																																																																																																																											
Risky habits	-0.053	(0.0324)	n.s.	-0.054	(0.0325)	n.s.	-0.100	(0.0375)	**	-0.119	(0.0378)	**																																																																																																																																																																																																																																																																																																											
Total risk	0.274	(0.0335)	***	0.281	(0.0337)	***	0.243	(0.0378)	***	0.269	(0.0381)	***																																																																																																																																																																																																																																																																																																											
Agriculture	-0.001	(0.0026)	n.s.	-0.011	(0.00263)	***	0.033	(0.00359)	***	0.023	(0.00366)	***																																																																																																																																																																																																																																																																																																											
Construction	-0.011	(0.00394)	**	-0.014	(0.00396)	**	-0.017	(0.0135)	n.s.	-0.021	(0.0137)	n.s.																																																																																																																																																																																																																																																																																																											
Trade&Horeca	0.002	(0.00342)	n.s.	0.001	(0.00345)	n.s.	0.018	(0.00309)	***	0.018	(0.00313)	***																																																																																																																																																																																																																																																																																																											
Businessservices	0.005	(0.00266)	n.s.	0.003	(0.00268)	n.s.	0.019	(0.00317)	***	0.018	(0.00321)	***																																																																																																																																																																																																																																																																																																											
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 (ref=Employees)													Self-employed	-0.030	(0.00788)	**	-0.030	(0.00788)	**	-0.036	(0.00884)	***	-0.356	(0.00884)	***	Agriculture				0.487	(0.0196)	***				0.515	(0.0263)	***	Industry				-0.396	(0.0162)	***				-0.137	(0.0243)	***	IRREGULAR JOB NACE (ref=HH services)				0.130	(0.0156)	***				-0.088	(0.0547)	n.s.	Trade&Horeca				-0.216	(0.0147)	***				-0.213	(0.0187)	***	Business services				-0.304	(0.0158)	***				-0.344	(0.0213)	***																																																																																																																																																																																																																													
Self-employed	-0.030	(0.00788)	**	-0.030	(0.00788)	**	-0.036	(0.00884)	***	-0.356	(0.00884)	***																																																																																																																																																																																																																																																																																																											
Agriculture				0.487	(0.0196)	***				0.515	(0.0263)	***																																																																																																																																																																																																																																																																																																											
Industry				-0.396	(0.0162)	***				-0.137	(0.0243)	***																																																																																																																																																																																																																																																																																																											
IRREGULAR JOB NACE (ref=HH services)				0.130	(0.0156)	***				-0.088	(0.0547)	n.s.																																																																																																																																																																																																																																																																																																											
Trade&Horeca				-0.216	(0.0147)	***				-0.213	(0.0187)	***																																																																																																																																																																																																																																																																																																											
Business services				-0.304	(0.0158)	***				-0.344	(0.0213)	***																																																																																																																																																																																																																																																																																																											

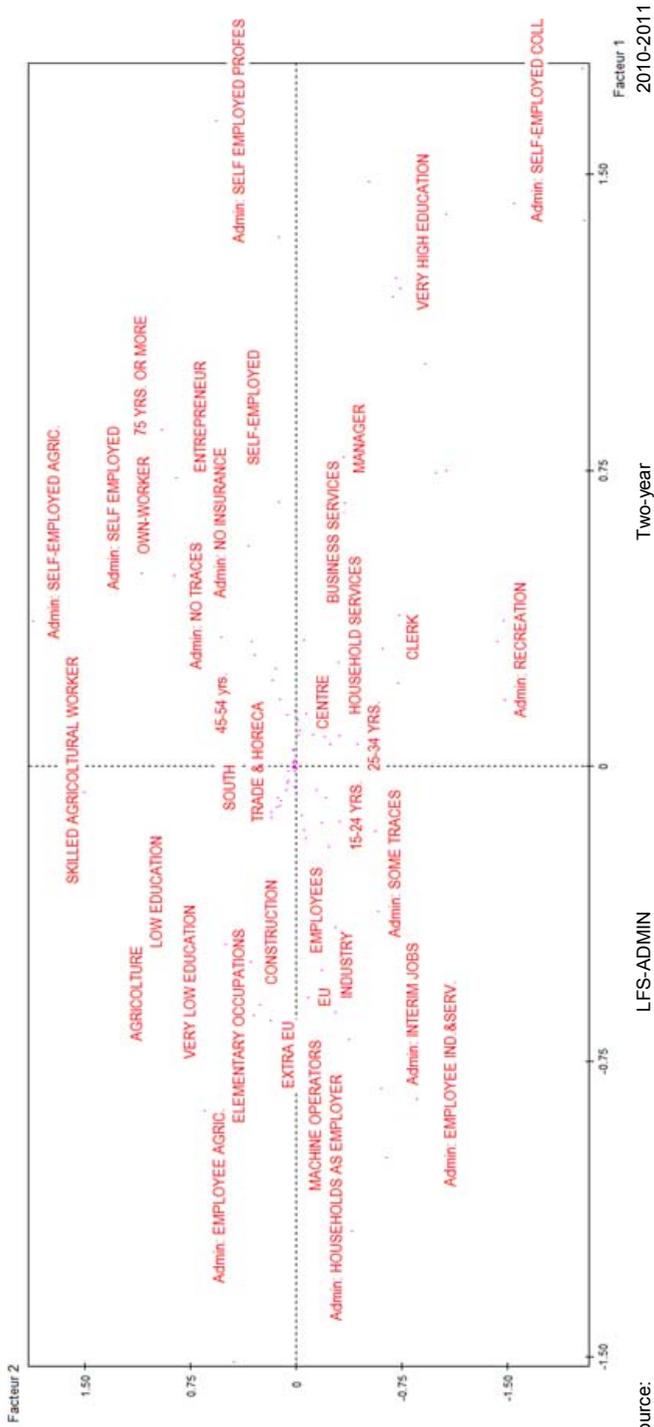
Note: *** if Pr. < 0.0001; ** if 0.0001 < Pr < 0.01; * if 0.01 < Pr < 0.05; * if 0.05 < Pr < 0.1; n.s. otherwise.

Table A.4. T-test on the gender differences in coefficients estimates by model (Men-Women)

Variables and modalities	MODEL					
	A		B		C	
Intercept	-10.1	**	2.7	**	2.7	**
CITIZENSHIP=Italian	-0.1	n.s.	-3.7	**	-1.7	n.s.
AGE=15-24	-3.9	**	-5.1	**	-8.7	**
AGE=25-34	-0.8	n.s.	-2.2	*	-5.3	**
AGE=55-64	2.6	**	2.7	**	1.8	n.s.
AGE=65-74	2.3	*	2.9	**	6.2	**
HOUSEHOLD=Single	8.7	**	8.5	**	8.4	**
HOUSEHOLD=Spouse(2Parents&1son)	-7.4	**	-8.3	**	-7.7	**
HOUSEHOLD=Spouse(2Parents&≥2sons)	-10.6	**	-12.3	**	-10.6	**
HOUSEHOLD=Spouse(2no sons)	-5.5	**	-5.8	**	-4.7	**
HOUSEHOLD=Spouse(2with other components)	-3.7	**	-3.8	**	-3.4	**
HOUSEHOLD=Son(1Parent&1son)	2.2	n.s.	3.7	**	3.4	**
HOUSEHOLD=Son(1Parent&≥2sons)	3.8	**	3.8	**	3.3	**
EDUCATION=ISCED 0-1	-6.0	**	-5.6	**	-5.8	**
EDUCATION=ISCED 2	-1.5	n.s.	-2.5	*	-2.6	*
EDUCATION=ISCED 3-4	-3.1	**	-2.7	**	-2.7	**
EDUCATION=ISCED 5	9.6	**	9.3	**	8.8	**
OTHER HOUSEHOLD INCOMES=Regular	-0.3	n.s.	1.6	n.s.	2.0	*
LABOURMARKET(bygender)=Factor2(b)			-2.0	*	-2.2	*
LABOURMARKET(bygender)=Factor3(c)			-2.4	*	-2.8	**
REGULAR EMPL.STRUCT.(bygender)=Agriculture			-7.7	**	-7.5	**
REGULAR EMPL.STRUCT.(bygender)=Trade&Horeca			-3.4	**	-3.6	**
REGULAR EMPL.STRUCT.(bygender)=Businessservices			-3.3	**	-3.5	**
EMPLOYMENT IN MICROENTERPRISES			2.2	*	2.4	*
TYPE OF JOB=Self-employed					27.5	**
IRREGULAR JOB NACE=Industry					-8.9	**
IRREGULAR JOB NACE=Construction					3.8	**

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

Chart A.1. Modalities on the first two factor's space



Source:

