VOLUME LXVIII – N. 3/4 LUGLIO-DICEMBRE 2014

RIVISTA ITALIANA DI ECONOMIA DEMOGRAFIA E STATISTICA

Sede Legale
C/O Studio Associato Cadoni, Via Ravenna n.34 – 00161 ROMA

sieds.new@gmail.com
rivista.sieds@gmail.com

Volume pubblicato con il contributo
dell'Istituto di Studi sulle Società del Mediterraneo del CNR di Napoli
e del Dipartimento di Scienze Politiche dell'Università Federico II di Napoli.

IN THIS ISSUE

Questo volume accoglie una selezione delle comunicazioni dei Soci presentate in occasione della 51esima Riunione Scientifica della Società Italiana di Economia, Demografia e Statistica. La Riunione Scientifica è stata organizzata in collaborazione con il Dipartimento di Scienze Politiche, Università Federico II, e con l'Istituto di Studi sulle Società del Mediterraneo del CNR-Napoli.
Un sentito ringraziamento va ai referee per l'accuratezza e l'importanza del lavoro svolto.

Claudio Ceccarelli

INDICE

Stefania Maria Lorenza Rimoldi, Elisa Barbiano di Belgiojoso
Detecting the poor among foreigners: remarks on a convenient equivalence scale 7

Luca Salvati, Marco Zitti, Giuseppe Venanzoni, Margherita Carlucci
Una nuova fotografia del divario tra Nord e Sud: disparità regionali degli indicatori socio-economici e ambientali 15

Claudio Ceccarelli, Giovanni Maria Giorgi, Alessio Guandalini
Is Italy a melting pot? 23

Anna Di Bartolomeo, Giuseppe Gabrielli, Salvatore Strozza
Policies and measures of integration in Italy: the cases of Moroccans and Ukrainians 31

Michele Lalla, Elena Pirani
The secondary education choices of immigrants and non-immigrants in Italy 39

Alessio Buonomo, Elena de Filippo, Giuseppe Gabrielli
Individual and household characteristics and migratory models of immigrants in Campania 47

Francesca De Palma, Stefania Girone, Sara Grubanov-Bošković
Looking back to look forward: the Italian active ageing in between the old and the new millennium 55

Luciano Nieddu, Cecilia Vitiello
Cluster weighted beta regression 63

Domenica Quartuccio, Giorgia Capacci
Povertà ed esclusione sociale delle famiglie in Italia 71

Antonella Bernardini, Andrea Fasulo, Marco D. Terribili
A model based categorisation of the Italian municipalities based on non-response propensity in the 2011 Census 79

Margherita Gerolimetto, Stefano Magrini
Spatial analysis of employment multilpliers in Spanish labor markets 87

Anna Di Bartolomeo, Salvatore Strozza
Immigrants living in the EU15 countries and their conditions of integration in the labour market 95

Agostino Di Ciaccio, Giovanni Maria Giorgi
Machine learning and text mining to classify tweets on a political leader 103

Silvia Loriga, Andrea Spizzichino
Le ore lavorate: un'analisi dei risultati della rilevazione sulle Forze Lavoro 111

Rosa Calamo, Thaís García Pereiro
Occupazione femminile: l'Olanda un esempio virtuoso per l'Italia? 119

Matteo Mazziotta, Adriano Pareto
A composite index for measuring Italian regions' development over time 127

Chiara Gigliarano, Francesco Maria Chelli
A nonparametric Gini concentration test for labour market analysis 135

Barbara Zagaglia, Eros Moretti
Fertility dynamics in Europe: reflections on the principal interpretative paradigms in light of some empirical evidence 143

Anna Maria Altavilla, Angelo Mazza, Luisa Monaco
Effetti dell'invecchiamento della popolazione sulla spesa del Sistema Sanitario Nazionale 151

Anna Maria Altavilla, Angelo Mazza, Antonio Punzo
A comparison of bias correction methods for the dissimilarity index 159

Gianni Bergamo, Claudio Pizzi
Foreign direct investment and psychic distance: a gravity model approach 167

Francesca Lariccia, Antonella Pinnelli, Sabrina Prati, Marina Attili, Claudia Iaccarino
L'appropriatezza del taglio cesareo nelle regioni italiane: analisi con la classificazione di Robson 175

Federico Benassi, Fabio Lipizzi, Donatella Zindato
Un'analisi geografica sulla presenza dei cittadini stranieri a Roma 183

Leonardo Di Marco, Luciano Nieddu
Trigger factors that influence bankruptcy: a comparative and exploratory study 191

Antonio Cappiello
Luigi Bodio: promoter of the political and high scientific mission of statistics and pioneer of the international statistical cooperation 199

Rivista Italiana di Economia Demografia e Statistica Volume LXVIII n. 3/4 Luglio-Dicembre 2014

DETECTING THE POOR AMONG FOREIGNERS: REMARKS ON A CONVENIENT EQUIVALENCE SCALE[1]

Stefania Maria Lorenza Rimoldi, Elisa Barbiano di Belgiojoso

1. Introduction

That foreign immigrants are more vulnerable to poverty than natives is a well evident fact in reality beyond scientific research, rich of contributions in this field (Lelkes, 2007; Kazemipur and Halli, 2011; Dalla Zuanna 2013, among others). Newspapers daily illustrate situations of social marginality sometimes so extreme to border on degradation of entire neighbourhoods, usually in the periphery of urban centres. Many organizations working in the third sector (Caritas, Banco Alimentare, Società San Vincenzo, Frati Francescani, etc.) document a chronic poverty among immigrants, even increased in recent years due to the economic juncture Italy is being experiencing (Rimoldi and Accolla, 2010; Blangiardo and Rimoldi, 2013). However, whatever its perception, a problem of measuring the incidence of poverty among immigrants arises when making use of tools designed for a population quite different, the Italian one. The discussion about the validity of the measurement tools involves the discussion about the different households' ability to convert resources into wellbeing, that means to ascertain whether the Carbonaro equivalence scale, conceived (thirty years ago) for Italian families may be valid also for foreign families.

2. Theoretical framework

Migrants move in search of opportunities that are not available in their country. At the beginning they are minded to accept a certain risk of experiencing a transitional period in poverty compared to natives, in the perspective of a global improvement of conditions compared to their countrymen who don't move. Then, immigrants can feel poor compared to natives but they feel non-poor compared

[1] Paragraphs 1-3 are due to Rimoldi S.M.L., paragraphs 4-5 are due to Barbiano di Belgiojoso E.

with their countrymen. It follows that poverty is a relative concept: the reference standard for the same individuals may be different. Therefore, subjective perception of poverty by immigrant can be described not as a dichotomous variable (poor and non-poor), but along a continuum of states ranging from the level of the country of origin (very poor) to the one of the country of destination (rich), acquired as a reference. The assessment of own poverty status determines the consumption behaviour, i.e, the ability to transform the available resources into well-being. It follows that the consumption behaviour (both in terms of quantity and quality of goods) of more integrated immigrants is more similar to the natives' one while significant differences are observed with respect to the less integrated immigrants. These gaps must be ascribed to at least two reasons. First, the immigrants' exceptional mobility (the higher the shorter the duration of presence) affects the size and shape of families. Immigrant families expand and shrink continuously to receive relatives or simply compatriots just arrived and the traditionally model "couple with children" is the goal to be reached in the long run. Second, differences in the standard of reference between country of origin and country of destination affect the economies of scale of families. It should also be noted that simple subsistence lifestyle is fairly common among immigrants, and forms of solidarity can exist between members of certain social groups where friends and relatives help families by providing them with even considerable quantity of consumer goods. Therefore, it seems evident that the consumption behaviour of immigrant families cannot, a priori, be measured with the same equivalence scale of the natives' families. There would be a coincidence between the two scales only in case of perfect integration and absence of frictional phenomena related to migration. It has been argued that "these problems of equivalence are important, but mainly only so far as they affect the precision of the estimate and not because they affect the fundamental conception of this approach to poverty measurement" (Greeley, 1994). We would suggest, on the other hand, that they are in fact conceptual problems, since poverty estimate is based on unshared standards of living and different consumption profiles among households. Economies of scale can play a determinant role in poverty analysis: failure to correctly identify household composition can therefore lead to biases in poverty results (Galloway and Aaberge, 2003).

3. Data and methods

The research issue materializes in building a specific equivalence scale for the immigrant families and in measuring the impact on the incidence of poverty.

The equivalence scale suggested hereafter refers to Engel's law according to which, as income rises, the proportion of income spent on food falls. The equivalence coefficients are computed by the ratios between the incomes of families of different size and composition, which spend the same income share for food, and are hence assumed to have the same living standard.

Waves 2004-2012 of the ORIM (Lombardy Region Observatory on Immigration) surveys are employed to estimate the so-called "foreign scale". Unfortunately, the average monthly total family expense is available only split into four categories: "food, clothes", "dwelling", "transport, leisure, instalments" and "remittances". We opted for a subjective approach for the respondents to indicate the primary goods in the first category. We also excluded housing costs that, especially in the early stages of the migration process, represent a minimal share of total expenditure: in these phases immigrants often share housing poor, overcrowded and poor quality. A final consideration refers to the exclusion of remittances in total expenditure: based on data, no univocal relationship can be detected between remittances and total expense, since remittances decrease even when total expense increases, therefore we decided not to take them into account. All the items have been deflated annual (NIC) in order to obtain monetary values at constant prices.

The interval of the observations 2004-2012 has been divided into three three-year periods, for a total of 51,695 cases.

Therefore, with X_h and $C_{a,h}$ being, respectively, the total and "food, clothes" expenditure for each h family, and n_h its size, the regression model can be written as follows (Vernizzi and Siletti, 2004):

$$\log C_{a,h} = \alpha + \beta \cdot \log X + \eta \cdot \log n_h .$$

Despite the limits highlighted by previous studies (e.g. Lemmi et al. 2014), in order to evaluate poverty among foreigners living in Italy, we adopted the *International Standard of Poverty Line* method since most national institutes of statistics adopt this method. This methodology is grounded on the estimate of a relative poverty line as an explicit function of the family income (or consumption expenditure), namely a constant fraction of some family income (or consumption expenditure) standard. We opted for income as the welfare indicator since the consumption expenditure of foreigners is strongly affected by migrants' behaviour characterised by the maximisation of savings and frequent remittances to their country of origin (Barbiano di Belgiojoso et al., 2009; Barsotti and Moretti, 2004). We took the mean per capita income as the threshold, as Banca d'Italia (2006,

2008, 2010, 2012) does. Hence, a two member household is considered poor if its family income is lower than the mean national per capita income. The income of different size households is made equivalent to that of a family of two members using both the Carbonaro scale and the foreign scale (Table 1). As our aggregation method, we opted for the headcount ratio. The incidence of poverty is computed on ORIM data 2007-2012 and on EU-Silc 2009, Italian foreign module[2].

4. Results

There are more economies of scale among foreign households than in Italian households[3] (Table 1). In order to keep the same level of wellbeing as a household with two components, foreign households with three or more members have to increase their income by a lower proportion compared to the Italian households. Migrants living alone, on the other hand, have a higher coefficient of equivalence. Thus, we postulate to find lower poverty incidence among the households with more members, which are usually more penalized by the Carbonaro scale.

Table 1 - *Coefficient of the equivalence scale by household size: Carbonaro and Foreign scale*

scale	Household size						
	1	2	3	4	5	6	7+
Carbonaro	0.59	1	1.34	1.63	1.91	2.15	2.40
Foreign	0.71	1	1.22	1.41	1.57	1.72	1.86

Source: authors' elaborations on ORIM data.

Using different equivalence scales leads to different incidence of poverty among foreign families (Table 2). More specifically, according to the scale here presented, the incidence of poverty is lower than in the case of the Carbonaro scale.

According to the ORIM data, the gap between the two estimates of poverty incidence is 5-7 percentage points, furthermore the gap increases over time. Based on Eu-Silc data, difference is only 1.7%, but it must be noticed how the sample population is distorted being affected by an overestimation of "singles", as widely documented by the 2001 Census data.

[2] With regards this source of data only foreigners from high emigration countries are considered.

[3] With the term "Italian" we refer to the set of households the Carbonaro scale is based on, that is, all the households living in Italy in the early 1980s. Notice that at that time immigration was far from being the sizeable phenomenon it is today, so the term Italian seems appropriate.

Table 2 - *Incidence of poverty among foreign families according to both Carbonaro and Foreign scale.*

ORIM						
Incidence of poverty	2007	2008	2009	2010	2011	2012
Foreign scale	24.1%	25.3%	27.4%	29.2%	29.1%	32.2%
Carbonaro scale	29.5%	29.2%	32.3%	34.9%	34.2%	39.0%

	EU-Silc	
	Carbonaro scale	Foreign scale
Not at risk of poverty	50,7%	52.2%
At risk of poverty	49.3%	47.8%

Source: authors' elaborations on ORIM data 2007-2012 and Eurostat EU-Silc 2009.

Some interesting findings emerge when comparing the different groups of poor according to the two equivalence scales. Special attention is paid to families when they are classified in different manner by the two scales. How many are they? Why are they "poor" for one scale and "non-poor" for the other? What characteristics do these families have?

Table 3 - *Distribution of foreign households according to Carbonaro and Foreign scale*.

Carbonaro scale	ORIM	
	Foreign scale (row percentages)	
	Non poor	Poor
Non poor	97.3%	2.7%
Poor	21.4%	78.6%
Carbonaro scale	EU-Silc	
	Foreign scale (row percentages)	
	Non poor	Poor
Non poor	93.9%	6.1%
Poor	9.3%	90.7%

Source: authors' elaborations on ORIM data 2007-2012 and Eurostat EU-Silc 2009.

Based on ORIM data in Table 3, there is a large number of families who are classified as "poor" according to the Carbonaro scale but who appear "non-poor" according to the foreign scale (henceforth referred as *PoC*, "poor only for Carbonaro"): as many as 21.4% (more than 1 in 5) of families classified as poor with the Carbonaro scale is classified differently according to the equivalence scale suggested here. As a consequence, the share of "poor" for both the scales (*AP*, "always poor") is 78.6%. As regards the "non-poor", there is no significant difference between the scales (in 97.3% of cases, hereafter named the *NP*, "never

poor", scales agree). Anyway, 2.7% of the "non-poor" for Carbonaro are classified as "poor" (*PoF*, "poor only for foreign scale") only for the foreign scale.

Eu-Silc data show for both *PoC* and *PoF* an incidence of about 6-9%, consistent with the hypothesis of an overestimation of singles in the sample.

Table 4 - *Main characteristics of foreign families according to the cross classification of the Carbonaro and Foreign scale.*

	always poor	poor only Carbonaro	poor only foreign scale	never poor
Household size in Italy (mean)	3.3	4.5	1.0	2.4
n. children (mean)	1.6	2.0	0.8	1.1
n. children in Italy (mean)	1.3	1.9	0.0	0.7
n. children abroad (mean)	0.3	0.1	0.8	0.5
living arrangement		80.7% live with partner/spouse with children	36.3% alone 73.7% with friends, relatives or acquaintances	
% home-ownership	15.2%	29.8%	2.6%	24.2%
% employed*	49.0%	62.4%	70.0%	81.3%
Duration of presence (mean)[a]	8.5	10.7	5.5	9.1
number of families	10,258	2,799	720	26,036

Note: (a) information available only for the interviewee considered as reference person of the family
Source: authors' elaborations on ORIM data 2007-2012.

Regardless of the dataset used (EU-Silc or ORIM) or the period (2007-2012) considered, the results of the analysis show a clear pattern in the cross-classified families. Actually, families who are classified as "poor" only according to one of the two compared equivalence scales (Carbonaro or foreign) have a precise socio-demographic profile (Table 4). More specifically, people classified as *PoC* are usually foreigners living in Italy with their household, more frequently as a couple with children and with or without other members. Moreover, they are typically homeowners, with a higher number of years since migration, and in the main workers with a long-term contract. Such a result seems surprising since all these features seem to indicate advanced settlement behavior, generally corresponding to a higher level of socio-economic integration than that of the AP group (Borjas, 2002, before others). Being a homeowner is usually strongly associated with being "non-poor" (e.g. Painter et al., 2001): the share of homeowners among PoC is

29.8% of families, versus 24.2% among NP. Moreover, we may consider the presence of the household as a sign of a higher standard of wellbeing in itself, since several conditions must be fulfilled in order to achieve family reunification (a regular permit of stay, a minimum size of accommodation and a minimum income, depending on the number of members to be reunified).
Whereas PoF are frequently present in Italy without their families, they are usually hosted by friends or by the community network, or they live at their workplace. Generally, they have just arrived in Italy, are often without a regular permit of stay, and they are employed in casual and seasonal jobs. Moreover, they frequently have no family left behind (neither spouse nor children at home).

5. Conclusions

In this study, we discussed the use of Carbonaro equivalence scale to estimate the level of poverty among foreigners. The results highlighted some significant elements that can contribute to the debate on the measurement of poverty among foreigners. In summary, the economies of scale between foreign families are higher than the Italian ones. By adopting a specific equivalence scale for foreigners a lower incidence of poverty is obtained as a first result. In addition, some important differences emerged with reference to the qualitative characteristics of the poor. In particular, the poor only for Carbonaro are families who have attained a high degree of social and economic integration. It follows that the Carbonaro scale would seem to overestimate the poverty of the families of foreigners just because are numerous. Well aware that our analyses (which are based, among other things, on limited data) do not solve the problem of defining "the" measure of poverty among foreigners, anyway we suggest that the introduction of a specific equivalence scale that takes into account the different economies (or diseconomies) of scale in foreign households calls attention to the consequences that ignore them entails. The analyses presented here indicate the need for further study on the basis of more detailed data on the consumption behaviour of foreign families (currently not available), also investigating specific population subgroups.

References

BANCA D'ITALIA 2008-2014. I bilanci delle famiglie italiane nell'anno 200.., *Supplementi al Bollettino Statistico*. Indagini campionarie, Nuova serie.

BARBIANO DI BELGIOJOSO, E., CHELLI, F.M., AND PATERNO, A. 2009. Povertà e standard di vita della popolazione straniera in Lombardia, *Rivista Italiana di Economia Demografia e Statistica* LXIII, 3/4, 23-30.

BARSOTTI, O., MORETTI, E. (eds.) 2004. *Rimesse e cooperazione allo sviluppo*, Franco Angeli, Milano.

BLANGIARDO G.C, RIMOLDI S.M.L. 2013. Atlante statistico della povertà materiale. In: Eupolis Lombardia (Ed.). *L'esclusione sociale in Lombardia: quarto rapporto 2011, 15-36*, Milano, Eupolis Lombardia.

BORJAS, G.J. 2002. Homeownership in the immigrant population, *Journal of Urban Economics*, 52, 448-476.

DALLA ZUANNA, G. 2013. Verso l'Italia, un modello di immigrazione. *Il Mulino*, 62, 1, 47-54.

GALLOWAY, T.A., AABERGE, R. 2003. Assimilation Effects on Poverty Among Immigrants in Norway. *MEMORANDUM, 07/2003*. Department of Economics University of Oslo.

GREELEY M. (1994). Measurement of Poverty and Poverty of Measurement. *IDS Bulletin.* 25.2. Institute of Development Studies.

KAZEMIPUR, A., AND HALLI, S. S. 2001. Immigrants and 'New Poverty': The Case of Canada1. *International Migration Review* 35, 4, 1129-1156.

LELKES, O. 2007. *Poverty Among Migrants in Europe*. Policy Brief April 2007. European Centre for Social Welfare Policy and Research.

LEMMI ET AL. 2014, Povertà e deprivazione. In: Saraceno, C., Sartor, N. and Sciortino, G. (Eds.): *Stranieri e diseguali. Le disuguaglianze nei diritti e nelle condizioni di vita degli immigrati*. Il Mulino, Bologna.

PAINTER, G., GABRIEL, S., AND MYERS, D., 2001. Race, Immigrant Status, and Housing Tenure Choice. *Journal of Urban Economics*, 49, 150-167.

RIMOLDI S.M.L., ACCOLLA G. 2010. La povertà in Lombardia attraverso i dati dell'osservatorio regionale sull'esclusione sociale. *Rivista Italiana di Economia, Demografia e Statistica*, 63, 179-186.

VERNIZZI, A., SILETTI, E. 2004. Estimating the cost of children through Engel curves by different good aggregates. *Statistical and Mathematical Applications in Economics*1036, 313-336

SUMMARY

A problem of measuring the incidence of poverty among immigrants arises when making use of tools conceived for the Italian population. In this study, we discuss the use of Carbonaro equivalence scale to estimate the poor among foreigners. The results highlight the need for a specific equivalence scale that takes into account the different economies of scale in foreign households.

Stefania RIMOLDI, Università Milano Bicocca, stefania.rimoldi@unimib.it
Elisa BARBIANO DI BELGIOJOSO, Università Milano Bicocca, elisa.barbiano@unimib.it

Rivista Italiana di Economia Demografia e Statistica Volume LXVIII n. 3/4 Luglio-Dicembre 2014

UNA NUOVA FOTOGRAFIA DEL DIVARIO TRA NORD E SUD: DISPARITÀ REGIONALI DEGLI INDICATORI SOCIO-ECONOMICI E AMBIENTALI

Luca Salvati, Marco Zitti, Giuseppe Venanzoni, Margherita Carlucci

1. Introduzione

Sviluppo sostenibile e coesione territoriale rappresentano due elementi chiave delle strategie nazionali ed europee. L'analisi della complessità dei fenomeni ambientali e delle loro interazioni con i processi socio-economici a livello locale rappresenta, quindi, non solo una sfida interpretativa per gli studiosi, ma anche - o soprattutto - un elemento cruciale di informazione da fornire al decisore politico per l'implementazione ed il monitoraggio di adeguate politiche di sviluppo regionale. L'occorrenza simultanea di degrado ambientale, segregazione sociale e polarizzazione economica, accelera i fenomeni di squilibrio territoriale ed è in grado di innestare una spirale perversa di conflitti sociali che mina alla base le possibilità di sviluppo sostenibile di intere regioni (Iosifides e Politidis, 2005; Kok et al., 2004; Onate e Peco, 2005). Una distribuzione sbilanciata delle risorse naturali ed economiche caratterizza in particolare i paesi europei del Mediterraneo (Zuindeau, 2007), per i quali l'impostazione di adeguati strumenti di policy richiede un approccio multidimensionale basato sull'analisi a livello locale delle interazioni tra fattori sociali, economici e ambientali (Puigdefabregas e Mendizabal, 1998; Salvati et al., 2014; Zuindeau, 2006).

Questo studio propone un'analisi integrata dei divari economici, ambientali e di sviluppo sostenibile a livello territoriale, con l'obiettivo di contribuire a delineare un quadro il più possibile completo dei legami spaziali tra le dinamiche economiche ed ambientali ed i sentieri di sviluppo (in)sostenibile osservati a livello locale. A tal fine viene confrontata la distribuzione per comune del principale indicatore di performance economica, il valore aggiunto pro capite, di un indicatore di qualità del capitale naturale, l'ESAI (Environmentally Sensitive Area Index), e di un indice di sviluppo sostenibile recentemente proposto per l'Italia (Salvati e Carlucci, 2014). I risultati dello studio intendono fornire indicazioni utili per l'implementazione di politiche tese al raggiungimento di uno sviluppo sostenibile

territorialmente bilanciato in paesi sviluppati che però, come l'Italia, presentano un grado notevole di disparità interne.

2. Metodologia

2.1. Caratteristiche degli indicatori utilizzati

L'analisi è stata condotta su tre indicatori, disponibili a livello comunale: a) un indicatore economico puro, il valore aggiunto pro capite, come *proxy* del livello di sviluppo economico e della competitività territoriale di ciascun comune, pubblicato dal Censis (2004) con riferimento temporale al 2001; b) un indicatore ambientale puro, l'ESAI, calcolato per il 2000; c) un indicatore composito di sviluppo sostenibile, che integra informazioni relative a tutti e tre i "pilastri" della sostenibilità, economico, sociale e ambientale, riferito all'anno 2001.

La metodologia ESAI (Environmentally Sensitive Area Index) è stata sviluppata nell'ambito del progetto europeo MEDALUS (MEditerranean Desertification And Land USe - DGXII, Ambiente) per l'individuazione di "aree sensibili dal punto di vista ambientale", attraverso un approccio basato su quattro fattori (suolo, clima, vegetazione e gestione del territorio) cruciali per la definizione del livello di vulnerabilità, in termini di disponibilità delle risorse naturali e di degrado ambientale, nelle regioni del Mediterraneo (Basso et al., 2000). Ad ogni fattore è associato un set di indici elementari (4 per i suoli, 3 per il clima, 4 per la vegetazione e 3 per la gestione del territorio), cui vengono attribuiti valori compresi fra 1 (predisposizione al degrado più bassa) e 2 (predisposizione più alta): ad esempio, per la qualità climatica si considera la media delle precipitazioni piovose, aridità ed esposizione dei versanti; la media geometrica delle componenti fornisce l'indice specifico per fattore, mentre l'ESAI si calcola come media geometrica dei quattro indici specifici. Il metodo ESAI è stato sottoposto a verifica sul campo in diversi paesi mediterranei, Portogallo, Spagna, Italia e Grecia (cfr. tra gli altri, Lavado Contador et al., 2009; Symeonakis et al., 2014).

L'indice composito di sviluppo sostenibile a livello comunale, che assume livelli compresi tra 0 e 1, è stato costruito come sintesi di 99 variabili relative a 14 dimensioni (Struttura della popolazione, Caratteristiche territoriali e struttura urbana, Istruzione, Mercato del lavoro, Struttura economica, Specializzazione turistica, Reddito e ricchezza delle famiglie, Criminalità, Gestione delle acque, Conduzione agricola, Paesaggio rurale, Caratteristiche delle coltivazioni agrarie, Qualità e innovazione in agricoltura, Capitale umano in agricoltura), riconducibili a 5 temi generali (Demografia, Capitale umano, Sviluppo locale e competitività, Qualità della vita, Sviluppo rurale e ambiente). I pesi assegnati a ciascuna variabile

sono stati determinati in base ai risultati di un'analisi in componenti principali (cfr. Salvati e Carlucci, 2014 per la metodologia di costruzione e le analisi di sensitività).

Come indicato in precedenza, i tre indici si riferiscono agli anni a cavallo dei Censimenti 2000/2001. Non è stato finora possibile estendere l'analisi ai Censimenti del 2011 in quanto il piano di diffusione dei risultati non è stato ancora portato completamente a termine e d'altro canto, solo la base censuaria permette un'adeguata disponibilità di dati al dettaglio comunale.

2.2. Analisi statistica

Una prima analisi descrittiva è stata effettuata sulle medie ed i coefficienti di variazione – assunti come *proxy* attendibile delle disparità territoriali per questo tipo di indicatori (cfr. Salvati e Zitti, 2008) - dei valori comunali dei tre indici a tre diversi livelli di aggregazione spaziale: per le 3 ripartizioni Nord, Centro e Sud; per le 20 regioni; per le 103 province (secondo le delimitazioni amministrative del 2001). Operando su diversi domini spaziali, infatti, è possibile verificare la stabilità dei risultati al variare della scala di aggregazione utilizzata e quindi tenere, almeno indirettamente, sotto controllo il problema dell'unità areale modificabile, ovvero la possibilità che i risultati di un'analisi spaziale varino a seconda dei confini e dell'ampiezza delle aree analizzate. Un ulteriore controllo tramite i coefficienti di correlazione binaria di Pearson è stato effettuato per verificare che le medie e i coefficienti di variazione dei 3 indicatori non fossero influenzati dal numero e dall'ampiezza dei comuni in ciascuna regione o provincia (in tutti i confronti si è avuto $p > 0,05$).

Per un'indicazione sintetica delle disparità territoriali, sono state condotte due analisi in componenti principali (ACP) sulle variabili rappresentate dalle medie e dai coefficienti di variazione dei 3 indicatori sulle 20 regioni e, rispettivamente, le 103 province. In entrambi i casi per l'ACP è stata considerata la matrice di correlazione e una soglia per la scelta degli autovalori principali pari all'unità. La presenza di correlazioni significative tra le variabili è stata controllata tramite il test di Bartlett e la misura di Kaiser-Meyer-Olkin.

3. Risultati

Tutti e tre gli indicatori utilizzati mettono in evidenza un netto gradiente Nord-Sud con le regioni settentrionali che non solo mostrano livelli più elevati di reddito, ma anche una migliore qualità ambientale ed un maggior livello di sostenibilità (Tabella 1). Ciò appare in contrasto rispetto all'opinione diffusa che la migliore

performance economica del Nord si accompagni a peggiori condizioni sociali ed ambientali (vedi ad esempio, Floridi et al., 2011).

Tabella 1 – *Valori medi e coefficienti di variazione degli indicatori per ripartizione**
** Per esigenze di spazio i valori regionali e provinciali non sono qui riportati, ma sono a disposizione presso gli autori.*

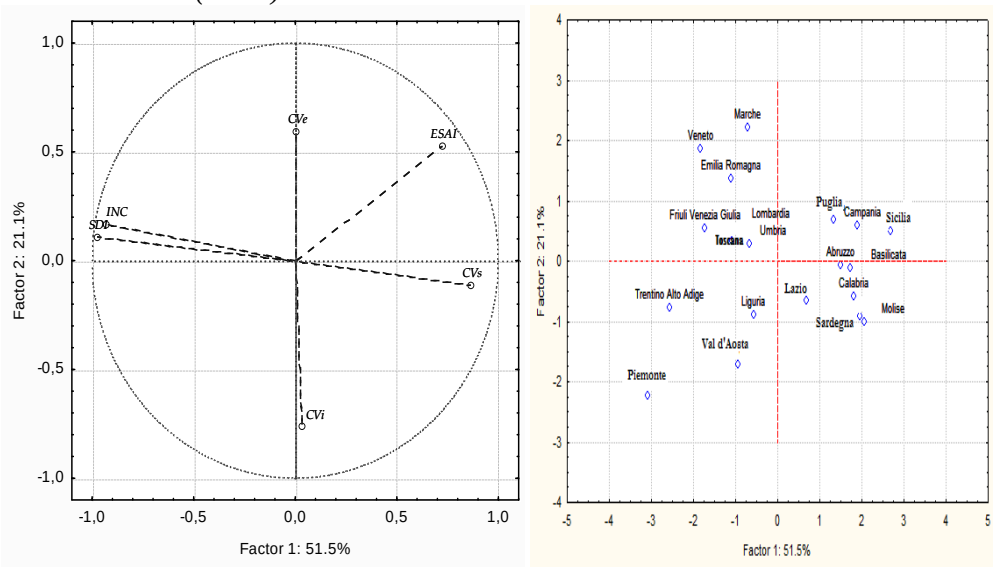
Ripartizione	Media	Coefficiente di variazione
	Valore aggiunto pro capite (medie in €)	
Nord	10,221	65
Centro	8,282	52
Sud	5,606	58
Italia	**8,549**	**69**
	ESAI (Environmentally Sensitive Area Index)	
Nord	1.338	5.6
Centro	1.353	5.3
Sud	1.398	4.6
Italia	**1.358**	**5.4**
	Indice di sviluppo sostenibile	
Nord	0.39	15
Centro	0.34	17
Sud	0.26	19
Italia	**0.34**	**23**

L'uniformità territoriale osservata nelle medie scompare quando si guarda alle disparità all'interno delle aree. La maggiore variabilità dei livelli di reddito si osserva al Nord, la più bassa al Centro, mentre i coefficienti di variazione degli indicatori di qualità ambientale e di sostenibilità mostrano gradienti opposti alla latitudine: per l'ESAI diminuisce scendendo dal Nord al Sud, per la sostenibilità invece aumenta. Questi andamenti sono confermati sia a scala regionale sia provinciale.

L'ACP a livello regionale ha estratto due componenti con autovalore superiore ad 1 ed una percentuale cumulata di varianza spiegata superiore al 72% (Figura 1). La prima componente (51% della varianza totale) è associata negativamente ai valori medi degli indicatori di reddito e di sostenibilità e positivamente con l'ESAI, confermando l'uniformità spaziale osservata nell'analisi descrittiva (valori alti dell'ESAI indicano una peggiore qualità ambientale). L'associazione positiva con il coefficiente di variazione dell'indice di sostenibilità sembrerebbe indicare che le regioni a più alto livello di sostenibilità siano anche più omogenee al loro interno. La seconda componente (21% della varianza totale) mostra chiaramente due andamenti territoriali contrapposti tra disparità ambientali ed economiche: le regioni con maggiori disparità di reddito mostrano una minore differenziazione interna nella qualità delle risorse naturali. Il coefficiente di variazione dell'indice di sviluppo sostenibile risulta incorrelato con gli altri due, suggerendo la possibilità di

meccanismi di compensazione fra differenze economiche ed ambientali (Munda e Saisana, 2011).

Figura 1 – *Risultati dell'ACP a livello regionale: pesi dei fattori (sinistra) e punteggi delle unità (destra)*



**** INC e CVi indicano media e coefficiente di variazione del valore aggiunto pro capite, SDI e CVs media e coeff. dell'indice di sviluppo sostenibile, CVe è il coeff. di variazione dell'ESAI***

I punteggi dell'ACP ordinano le regioni italiane lungo la prima componente, secondo il tipico gradiente Nord-Sud, mentre per la seconda componente le regioni del Nord e del Centro si dividono in due gruppi, in base alle disparità interne ambientali ed economiche. Anche l'ACP effettuata sulle province (2 componenti che spiegano complessivamente il 68% della varianza totale) conferma il gradiente Nord-Sud, ma in particolare identifica le province dell'Italia centrale come un'area caratterizzata da condizioni intermedie per tutti e tre i fenomeni analizzati.

4. Conclusioni

I risultati ottenuti suggeriscono come la distribuzione spaziale dei tre indicatori sia influenzata da una configurazione spaziale sfaccettata, con potenziali impatti sull'efficacia delle politiche locali di sviluppo regionale. L'Italia rappresenta un esempio di divisione territoriale determinata da dinamiche divergenti di fattori endogeni e di sentieri di sviluppo non pienamente sostenibili. In questo contesto, è

necessario che gli interventi di sviluppo siano tarati sulle specificità dei contesti regionali, attribuendo un ruolo cruciale ai nessi causali tra uno sviluppo territorialmente bilanciato ed i processi economici e ambientali.

La procedura qui presentata, applicabile anche ad altri contesti che presentino analoghe complessità economiche ed ambientali, appare, con alcuni *caveat* necessari, potenzialmente in grado di:

a) contribuire all'interpretazione di processi territoriali multidimensionali in termini interdisciplinari, tenendo comunque conto del fatto che i dati utilizzati - indicatori e 'medie' comunali – potrebbero già scontare in parte una riduzione delle dimensioni informative;
b) integrare dati provenienti da fonti differenti fornendo agli *stakeholders* locali strumenti, anche grafici, di interpretazione dei risultati ottenuti. idonei a contesti non accademici che necessitano di informazioni intuitive e immediate;
c) identificare un quadro rappresentativo delle disparità territoriali interne ad un'area come obiettivo per le politiche di sostenibilità.

Fornire strumenti informativi per le politiche territoriali di sviluppo sostenibile rappresenta un obiettivo particolarmente ambizioso in un'ottica multi-temporale. La struttura delle correlazioni individuata tramite l'ACP può modificarsi nel tempo, influenzando i contenuti informativi ed i risultati finali dell'analisi, ad esempio in termini di assi estratti e di varianza spiegata. Questo comporta una difficoltà intrinseca nel confronto tra risultati multivariati derivati da strutture di dati relative a due punti temporali distinti (ad es. due censimenti). A tal riguardo, tecniche di analisi *multi-way*, specificamente rivolte all'analisi del fattore tempo nell'ambito di un sistema di assi fattoriali rappresentati da un comparabile numero di variabili osservate sullo stesso supporto spaziale, possono rappresentare una soddisfacente soluzione analitica al problema (Salvati e Zitti, 2008).

Comprendere le complesse interazioni spaziali collegate agli aspetti economici ed ambientali per agire sulle disparità territoriali rappresenta, infatti, uno strumento importante per l'implementazione e il monitoraggio delle politiche nei paesi Mediterranei, ecologicamente fragili ed economicamente polarizzati (Nourry, 2008). Sviluppare un approccio simile a quello qui proposto, con la disponibilità di dati tempestivi e aggiornati regolarmente, potrebbe rivelarsi utile per determinare l'efficacia dei sentieri di sviluppo sostenibile intrapresi a livello locale.

Ringraziamenti

Il lavoro è stato finanziato con fondi di ricerca da Sapienza Università di Roma.

Riferimenti bibliografici

BASSO, F., BOVE, E., DUMONTET, S., FERRARA, A., PISANTE, M., QUARANTA, G., TABERNER, M. 2000. Evaluating environmental sensitivity at the basin scale through the use of geographic information systems and remotely sensed data: an example covering the Agri basin - Southern Italy, *Catena*, Vol. 40, pp. 19-35.

FLORIDI, M., PAGNI, S., FALORNI, S., LUZZATI, M. 2011. An exercise in composite indicators construction: Assessing the sustainability of Italian regions. *Ecological Economics*, Vol. 70, pp. 1440-1447.

IOSIFIDES, T. POLITIDIS, T. 2005. Socio-economic dynamics, local development and desertification in western Lesvos, Greece, *Local Environment*, Vol. 10, pp. 487-499.

KOK, K., ROTHMAN, D.S., PATEL, M. 2004. Multi-scale narratives from an IA perspective: Part I. European and Mediterranean scenario development, *Futures*, Vol. 38, pp. 261-284.

LAVADO CONTADOR, J.F., SCHNABEL, S., GOMEZ GUTIERREZ, A., PULIDO FERNANDEZ, M. 2009. Mapping sensitivity to land degradation in Extremadura, SW Spain, *Land Degradation and Development*, Vol. 20, pp. 129–44.

MUNDA, G., SAISANA, M. 2011. Methodological considerations on regional sustainability assessment based on multicriteria and sensitivity analysis. *Regional Studies*, Vol. 45, pp. 261-276.

NOURRY, M. 2008. Measuring sustainable development: some empirical evidence for France from eight alternative indicators. *Ecological Economics* Vol. 67, pp. 441-456.

ONATE, J.J., PECO, B. 2005. Policy impact on desertification: stakeholders' perceptions in southeast Spain, *Land Use Policy,* Vol. 22, pp. 103-114.

PUIGDEFABREGAS, J., MENDIZABAL, T. 1998. Perspectives on desertification: western Mediterranean, *Journal of Arid Environments*, Vol. 39, pp. 209-224.

SALVATI, L., CARLUCCI, M. 2014. A composite index of sustainable development at the local scale: Italy as a case study, *Ecological Indicators*, Vol. 43, pp. 162-171

SALVATI, L., ZITTI, M. 2008. Regional convergence of environmental variables: empirical evidences from land degradation. *Ecological Economics* Vol. 68, pp. 162-168

SALVATI, L., ZITTI, M., CARLUCCI, M. 2014. Territorial Systems, Regional Disparities and Sustainability: Economic Structure and Soil Degradation in Italy, *Sustainability*, Vol. 6, pp. 3086-3104.

SYMEONAKIS, E., KARATHANASIS, N. KOUKOULAS, S., PANAGOPOULOS, G. 2014. Monitoring sensitivity to land degradation and desertification with the Environmentally Sensitive Area Index: the case of Lesvos island, *Land Degradation & Development* (in press) DOI: 10.1002/ldr.2285.

ZUINDEAU, B. 2006. Spatial approach to sustainable development: Challenges of equity and efficacy, *Regional Studies*, Vol. 40, pp. 459–470.

ZUINDEAU, B. 2007. Territorial equity and sustainable development, *Environmental Values*, Vol. 16, pp. 253-268

SUMMARY

A new snapshot of the Italian North-South divide. Regional differences in socio-economic and environmental indicators

The study analyzes the distribution of per capita value added, a sustainable development index and an index of quality of the natural capital in Italy by municipality. A comparative analysis was carried out at three different spatial scales: (i) three geographical divisions, (ii) 20 administrative regions and (iii) 103 provinces. While the distribution of the three indicators was coherent across space, regional differences measured through the coefficient of variation for each of the three indicators showed totally decoupled patterns. On average, a high level in the sustainable development index corresponds to low regional disparities in the same index, while income and natural capital disparities were decoupled from the average level of the respective variables. On the whole, a marked north–south gradient reflecting the classical socioeconomic divide was observed between competitive and disadvantaged regions.

Luca SALVATI, Consiglio per la Ricerca e la Sperimentazione in Agricoltura (CRA-RPS), luca.salvati@entecra.it

Marco ZITTI, Consiglio per la Ricerca e la Sperimentazione in Agricoltura (CRA-CMA), marcozitti@libero.it

Giuseppe VENANZONI, Sapienza Università di Roma, Dipartimento di Scienze Sociali ed Economiche, giuseppe.venanzoni@uniroma1.it

Margherita CARLUCCI, Sapienza Università di Roma, Dipartimento di Scienze Sociali ed Economiche, margherita.carlucci@uniroma1.it

Rivista Italiana di Economia Demografia e Statistica Volume LXVIII n. 3/4 Luglio-Dicembre 2014

IS ITALY A MELTING POT?

Claudio Ceccarelli, Giovanni Maria Giorgi, Alessio Guandalini

1. Introduction

A melting pot is a metaphor for a society where many different types, mainly for ethnicity, race and consequently for culture, of people blend together as one. In an ideal situation it is a society in which these differences do not affect the social status of people. The United States is the classic example of a melting pot. However, there are other several examples in the world such as Afghanistan, Brazil and Israel.

Historically, Italy has always been an emigration country. Only since the seventies has started to become an immigration country. Earlier this shift to immigration was due to its economic situation and, later, mainly, for its position as the entry door of the Eurozone. Therefore, the migration problem and the migration policies are quite recent.

Nowadays, among the European countries, Italy ranks third for absolute number of foreign inhabitants (4.8 million) and eleventh for percentage of foreigners in the total population (5.5%). This work aims to evaluate the integration process of immigrants in Italy and see if our country can be considered a melting pot. Looking at the employee income, an ideal situation in which the foreign inhabitants can be considered integrated, at least for the employee wages, occurs if their incomes overlap with incomes of Italian inhabitants. On the contrary, we could state that the migration policies have been completely erroneous if the foreign inhabitants are the poorest whilst the Italians are the richest. That is, if the population is perfectly stratified.

The peculiarity of the work is represented by the tool used in evaluating the integration process and the migration policies, the analysis of Gini (ANOGI). The ANOGI is similar to the ANOVA (analysis of variance), but it offers an additional parameter: the stratification that enables us to better interpret the results. The work is more focused on the methodological aspects. In the first part, Section 2, the methodological differences between the ANOGI and the ANOVA are investigated. In Section 3, through the application on Italian Labour Force Survey 2007 and 2012 data the differences between the two methods are better clarified. Finally, an analysis of the integration process of immigrants is carried out.

2. Analysis of Gini (ANOGI) and analysis of variance (ANOVA)

2.1 ANOVA

The ANOVA is a well-known method to evaluate the differences between group means and their associated procedure. In the ANOVA setting, the observed variance in a particular variable is partitioned into components attributable to different sources of variation.

In the simplest case, the one-way ANOVA, the data Y_{ij} are assumed to be

$$Y_{ij} = \mu + a_i + \varepsilon_{ij}, \qquad i = 1, \dots, k \quad j = 1, \dots, n_i.$$

In this formulation the values Y_{ij} are expressed in function of a grand mean, μ, that is the common mean level of the treatment (or variable modality), and the unique effect due to treatment (or variable modality) a_i, besides the errors ε_{ij}.

The expected value of the errors are assumed to be independent and normally distributed with 0 mean and finite variance σ^2 equal for all the i (homoschedasticity). In formulas

i. $\mathbb{E}[\varepsilon_{ij}] = 0$;
ii. $\mathrm{Var}(\varepsilon_{ij}) = \sigma_i^2 < \infty$;
iii. $\sigma_i^2 = \sigma^2 \ \forall\, i$;
iv. $\mathrm{Cov}(\varepsilon_{ij}, \varepsilon_{i'j'}) = 0$ with $i \neq i'$ and $j \neq j'$;
v. $\varepsilon_{ij} \sim N(0, \sigma^2)$;

The basic idea of the ANOVA is that the variation is allocated to different sources. In fact, the overall variation of a measurable variable (left-hand side) is decomposed in two terms (right-hand side): between variation due only to treatments and within variation due only to random error, respectively. That is,

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(y_{ij} - \bar{y}\right)^2 = \sum_{i=1}^{k} n_i\,(\bar{y}_i - \bar{y})^2 + \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(y_{ij} - \bar{y}_i\right)^2$$

where $\bar{y}_i = \frac{1}{n_i}\sum_j y_{ij}$ and $\bar{y} = \sum_j n_i\, y_i / \sum_i n_i$. The corrected (by degree of freedom) sums of squares, under the ANOVA assumptions, are chi squared random variables. In particular, the left-hand side is distributed as a χ^2_{N-1} while, under the null hypothesis (equal means among the groups), the right-hand side is the sum of two independent random variables distributed, respectively, as χ^2_{k-1} and χ^2_{N-k}.

2.2 ANOGI

The ANOGI was firstly proposed by Frick et al. (2006). It is based on the Gini index that in a population P is defined as (Lerman and Yitzhaki, 1989, p. 44)

$$G = \frac{2\,\mathrm{cov}\left(y, F(y)\right)}{\mu},$$

that is, twice the covariance between the income y and the rank $F(y)$, standardized by mean income μ. When the population is divided in k groups, $P = P_1 \cup P_2 \cup \ldots \cup P_k$, the Gini index can be expressed as (Yitzhaki, 1994, p. 154)

$$G_u = \sum_{i=1}^{k} s_i G_i O_i + G_b, \qquad (1)$$

that is, the Gini index is decomposed in two components: within and between, where

i. $s_i = p_i\,\mu_i/\mu$ is the ratio between the mean of variable y in the group i, μ_i, weighted by its share, p_i, and the mean of y calculated on the whole population;
ii. G_i is the Gini index within group i;
iii. O_i is the overlapping index of group i with the entire population;
iv. G_b is the between-group inequality.

Two elements in (1) must be pointed out: overlapping and between-group inequality. Overlapping should be interpreted as the inverse of stratification (see, e.g., Yitzhaki, 1988, p. 39; Yitzhaki and Lerman, 1991, p. 319). It measures to what extent one group is overlapped by the other. The overlapping index O_i may be expressed as

$$O_i = \frac{\mathrm{cov}_i\left(y, F_u(y)\right)}{\mathrm{cov}_i\left(y, F_i(y)\right)},$$

that is the ratio between the covariance of y and the rank of units belonging to group i, calculated on their position in the overall distribution, and one-forth of Gini's mean difference of group i (see Yitzhaki and Schechtman, 2009, p. 149).

The overlapping index related to a given group i can be written in terms of the overlapping index between two groups, i and j,

$$O_i = \sum_j p_j\,O_{ij} = p_i O_{ii} + \sum_{j \neq i} p_j\,O_{ji} = p_i + \sum_{j \neq i} p_j\,O_{ji}$$

where

$$O_{ji} = \frac{\text{cov}_i\left(y, F_j(y)\right)}{\text{cov}_i\left(y, F_i(y)\right)}$$

represents the overlapping index of group j by group i (Yitzhaki, 1994). In particular:

i. $O_{ji} = 0$, when no member of group j lies in the range of subgroup i;
ii. $O_{ji} = 1$, the distributions of group i and j are identical;
iii. O_{ji} is not symmetrical, that is the higher O_{ji} the lower O_{ij};
iv. $O_{ji} \leq 2$; that is its maximum value, if all the members of group j are included between the members of group i and they are concentrated around the mean of group i.

The between group inequality

$$G_b = \frac{2\,\text{cov}\left(y, \bar{F}_{ui}(y)\right)}{\mu_u},$$

which is the ratio between twice the covariance between the mean of variable y of each group and the groups mean rank in the whole population and the mean of y.

When the population is perfectly stratified the between-group inequality is equal to the between-group-Pyatt inequality, (Pyatt, 1976, p. 247)

$$G_b^p = \frac{2\,\text{cov}\left(\mu_i, \bar{F}_i(y)\right)}{\mu_u}.$$

Yitzhaki and Lerman (1991, p. 322) demonstrated that $G_b^p \geq G_b$. In fact, G_b reaches its upper level as the overlapping index is equal to 0 and, therefore, the amount of total inequality is explained by the between inequality.

Introducing the between-group-Pyatt inequality, (1) can be written as

$$G_u = \sum_{i=1}^{k} s_i G_i + \sum_{i=1}^{k} s_i G_i (O_i - 1) + G_b^p + \left(G_b - G_b^p\right) = \\ = IG + IGO + BG + BGO \qquad (2)$$

that is, in terms of the four elements at the basis of ANOGI: the within (IG) and the between-group (BG) components and the effects of overlapping on within and between-group component, IGO and BGO, respectively.

2.3 Similarities and differences between ANOVA and ANOGI

The ANOVA and the ANOGI perform the same task; that is, they decompose a measure of variability, variance or Gini index respectively, and assign it to different sources of variation. Their components are conceptually comparable. As briefly illustrated in Table 1, IG has the same meaning as SSW in the ANOVA and the BG as SSB. In other words, both methods decompose the variability into two quantities: the difference within the groups and the difference between the groups.

Table 1 – *Comparison among components of ANOVA and ANOGI.*

	ANOVA		ANOGI	
Within	$SSW = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(y_{ij} - \bar{y}_i\right)^2$	$0 \leq SSW \leq SST$	$IG = \sum_{i=1}^{k} s_i G_i$	$0 \leq IG \leq G_u$
Between	$SSB = \sum_{i=1}^{k} n_i\,(\bar{y}_i - \bar{y})^2$	$0 \leq SSB \leq SST$	$BG = G_b^p$	$0 \leq BG \leq G_u$
Overlapping Within			$IGO = \sum_{i=1}^{k} s_i G_i (O_i - 1)$	
Overlapping Between			$BGO = \left(G_b - G_b^p\right)$	$-BG - IGO - IG \leq BGO \leq 0$

Moreover, to extra parameters linked to the overlapping, IGO and BGO, are derived with the ANOGI. IGO provides the contribution of each group to within group variability and tell us how much the distributions are intertwined and, therefore, how much the groups are integrated with one another. BGO is related to the effect of overlapping on the between-group inequality. It is always negative, because the overlapping reduce the ability to distinguish between groups.

3. The degree of melting pot

The advantage of the ANOGI with respect to the ANOVA is that it says how much a population is stratified and, on the contrary, how much the groups are intertwined. In this paper the ANOGI is used to investigate the integration of immigrants into the labour market in terms of employee wages. This paper traces out the work by Yitzhaki and Schechtman (2009).

From the Labour Force Survey 2007 and 2012 the employees older than thirty have been selected in order to avoid the effect of different fertility rates between Italians and immigrants. The employees have been split in three main categories, Italians, immigrants and second-generation immigrants, through the variables

gathered and in the questionnaire and in accordance with the Italian laws in matter regarding citizenship[1]. Furthermore the immigrants are also classified by geographical areas of origin (Europe, North-America, Center-America, South-America, Africa, North-Africa, Asia, Middle-East, China and Oceania).

The employees classified as second-generation immigrants in one case are aggregate to the Italians (wide classification, W) and, in another case, to the immigrants categorized by their geographical areas of origin (narrow classification, N). In both cases the ANOVA and the ANOGI are applied and the results obtained separately for each classification are compared to derive conclusions on the immigrants' integration.

3.1 ANOVA results

The ANOVA decomposes the total amount of variance in two quantities, between and within (Table 2).

Table 2 – *Results of the ANOVA analysis on Labour Force Survey data of 2007 and 2012.*

		MS between	MS within	Total (df)	SS Between (df)	SS within (df)	F
2007	N	55,640,686	312,283	45,691,375,078 (144,365)	612,047,549 (11)	45,079,327,529 (144,354)	178.17
	W	56,790,114	312,196	45,691,375,078 (144,365)	624,691,252 (11)	45,066,683,826 (144,354)	181.91
2012	N	142,358,333	324,943	44,166,251,741 (131,112)	142,358,333 (11)	42,600,310,082 (131,101)	438.10
	W	4,832,205	336,482	44,166,251,741 (131,112)	53,154,254 (11)	44,113,097487 (131,101)	14.36

Looking at the F ratio the MS between is larger for definition W than for N in 2007 while, in 2012 the contrary occurs. The evidence that the null hypothesis (equal means among the groups) must be rejected is stronger in these cases[2]. This means that in 2007, when the second-generation immigrants is classified as Italians a better stratification is performed while, in 2012, a better classification is reached when the second-generation immigrants is classified as foreigners.

[1] In the 2007's sample the employed were about 145 thousand representative of 12,7 millions in the population: 132 thousand were Italians, 7,5 thousand were immigrants and 4 thousand were second-generation immigrants, representative of 12.3, 0.9 and 0.4 millions of employed in the population, respectively.
In the 2012's sample the employed became about 131 thousand representative of 13,3 millions in the population: 113 thousand were Italians, 13,9 thousand were immigrants and 4,2 thousand were second-generation immigrants, representative of 12.3, 1.6 and 0.4 millions of employed in the population, respectively.

[2] Even considering the Welch's test (Welch, 1947) in the case of non-homogeneity of the variances the evidence is to reject the null hypothesis.

3.2 ANOGI results

Performing the ANOGI on the same data, it is possible to decompose the Gini index into Gini between-groups, Gini within-groups and overlapping. In 2007 the Gini between groups (G_b and also G_b^p) is larger for W – with respect to N – even if the values are close to one another. Instead, in 2012 the Gini between-groups is larger for N than for W. The overlapping index of N definition decreases from 2007 to 2012 whilst that of W definition increases and, therefore, the gap between the two indices becomes larger. This means that in 2007, when the second-generation immigrants are classified as Italians a better stratification is performed, whilst in 2012 a better classification is reached when the second-generation immigrants are classified as foreigners.

In all cases the larger part of the inequality is explained by the within groups inequality (SGO). The overlapping that affected the within inequality is negligible and almost all affects the between-groups inequality. Therefore, the ratio between G_b and G_b^p is crucial to evaluate the stratification of the employee wages. In 2007 a better stratification is obtained for definition W, whilst in 2012 for definition N. This means that in 2007 the second generation of immigrants had employee wages more similar to the Italians, but this is not true for 2012. Therefore, it is possible to state that the integration process had suffered a setback.

Table 3 – *Results of the ANOGI analysis on Labour Force Survey data of 2007 and 2012.*

	Overall Gini	Definition	SGO		G_b		G_b^p	G_b/G_b^p
2007	0.2214 (0.0008)	N	0.2153	97.27%	0.0061	2.73%	0.0172	0.355
		(SE)	(0.0008)		(0.0003)		(0.0005)	
		W	0.2151	97.18%	0.0062	2.82%	0.0152	0.408
		(SE)	(0.0009)		(0.0005)		(0.0002)	
2012	0.2258 (0.0008)	N	0.2137	94.64%	0.0121	5.36%	0.0302	0.401
		(SE)	(0.0008)		(0.0003)		(0.0005)	
		W	0.2254	99.84%	0.0004	0.16%	0.0025	0.160
		(SE)	(0.0008)		(0.0001)		(0.0002)	

4. Conclusion

The ANOVA and the ANOGI perform the same task, but the latter provides an extra parameter, the overlapping, that is useful to better interpret the results. The two methods have been applied to the employee wages from the Labour Force Survey of 2007 and 2010 to investigate the integration of immigrant in the Italian society and, in particular, the labour market but, moreover, to point out the similarities and differences between the two methods. Both the results of the ANOVA and of the ANOGI demonstrate that there was a step back in the

integration process from 2007 to 2012. Looking at the ANOGI results, it is possible to state that the second generation of immigrants was better integrated in 2007 than in 2012. However, in the global evaluation of the results it is important to point out that the application refers to employees with regular labour contract who have a higher level of integration in Italian society.

Acknowledgements

The present work has been realized within the grant for the project "Indici classici di disuguaglianza e variabilità: nuove prospettive di ricerca" (Sapienza 2013).

References

FRICK J.R., GOEBEL J., SCHECHTMAN E., WAGNER G.G., YITZHAKI S. (2006). Using Analysis of Gini (ANOGI) for Detecting Whether Two Sub-Sample Represent the Same Universe: The German Socio-Economic Panel Study (SOEP) Experience. *Sociological Methods & Research*, Vol. 34, No. 4, pp. 427-468.

LERMAN R.I., YITZHAKI S. (1984). A Note on the Calculation and Interpretation of the Gini Index. *Economics Letters*, Vol. 15, No. 3-4, pp. 363-368.

PYATT G. (1976).On the Interpretation and disaggregation of Gini Coefficient. *Economic Journal*, Vol. 86, No. 342, pp. 243-255.

YITZHAKI S. (1988). On Stratification and Inequality in Israel. *Bank of Israel Economic Review*, Vol. 63, No. 1-2, pp. 36-51.

YITZHAKI S., LERMAN R.I. (1991). Income Stratification and Income Inequality. *Review of Income and Wealth*, No. 37, No. 3, pp. 313-329.

YITZHAKI S., SCHECHTMAN R.I. (2009). The "Melting Pot": A Success Story?. *Journal of Economic Inequality*, Vol. 7, No. 2, pp. 137-151.

WELCH B.L. (1947). The generalization of Student's problem when several different population variance are involved. *Biometrika*, Vol. 34, No. 1-2, pp. 28-35.

SUMMARY

The immigrants integration process in Italy is investigated through the analysis of Gini (ANOGI). This methodology has an advantage with respect to the analysis of variance (ANOVA) because it provides a further element: the overlapping index, split in overllaping between and within the groups. This enables us to better understand and examine the immigrants integration looking at the stratification of the subpopulation of Italians and immigrants. The ANOGI is compared to the ANOVA and, then, the two methods are applied to Italian Labour Force Survey data of 2007 and 2012.

Claudio CECCARELLI, Italian National Institute of Statistics, Social and environmental statistics department, clceccar@istat.it.

Giovanni Maria GIORGI, "Sapienza" University of Rome, Department of Statistical Sciences, giovanni.giorgi@uniroma1.it.

Alessio GUANDALINI, "Sapienza" University of Rome, Department of Statistical Sciences, alessio.guandalini@uniroma1.it.

POLICIES AND MEASURES OF INTEGRATION IN ITALY: THE CASES OF MOROCCANS AND UKRAINIANS[1]

Anna Di Bartolomeo, Giuseppe Gabrielli, Salvatore Strozza

1. Introduction

Since the 1990s, scholars and policymakers have pointed out the necessity of studying migrants' integration within the different contexts of the Italian society. In recent years, such interest has grown in parallel with the gradual stabilization of the foreign presence in the country. The Turco-Napolitano Law (n. 40/1998) and the following Consolidated Law (Decreto legislativo n. 286 of 1998) established, for the first time in Italy, the Commission for Integration Policies of Immigrants. The Commission wished to move towards a "*reasonable integration*" model (Zincone, 2000). At that time, it was already clear that the process of integration - dynamic and multi-dimensional - necessarily involves a number of fields, namely social and cultural relations, the labor market, housing and living conditions, education and training, political rights and active citizenship.

The term integration therefore expresses a complex concept, whose meaning can vary in time and space (Golini, 2006). The same applies to the population of interest (Bonifazi, Strozza, 2003): in old destination countries, the challenge has long been to provide children and grandchildren of immigrants (second and third generations) with the same opportunities of autochthonous peers, by supporting their social mobility through education and adequate employment; in Italy, together with other European countries that have become new destination areas during last 20-30 years, scholars have long paid attention to first generation migrants (Cesareo, Blangiardo, 2009), while considering the school insertion of second generation migrants only during last decade (Dalla Zuanna et al., 2009).

This paper is part of a larger research project coordinated by the European University Institute - Migration Policy Centre and co-funded by the European Union (EU). The project analyses the integration of immigrants coming from Third Countries and residing in the EU27 by looking at integration as a process which

[1] This work is the result of a close collaboration between the authors. As for this version, paragraph 1has been written by the three authors; paragraph 5 by A. Di Bartolomeo; paragraphs 3 and 4 by G. Gabrielli; paragraph 2 by S. Strozza.

involves three main actors: the immigrant, the origin country and the destination country.

This contribute focuses on two national groups of immigrants, which are quantitatively important in the Italian case and are very different for demographic characteristics, migration patterns and insertion modalities: Moroccans and Ukrainians.

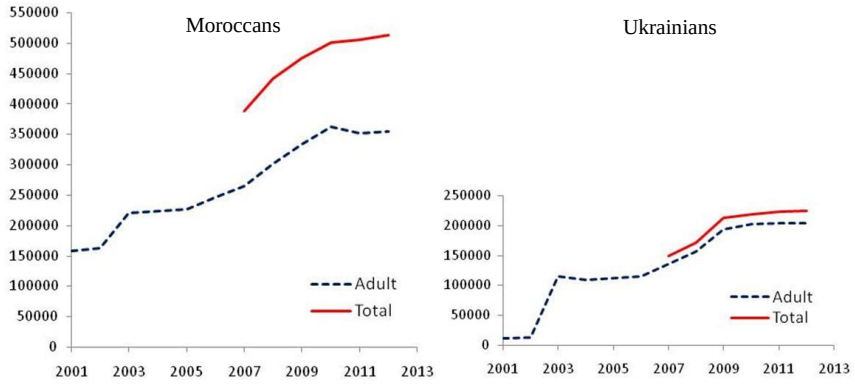
After a synthetic overview of the migratory evolution and the main demographic characteristics of the two observed groups (par. 2), we describe used data and methods (par. 3) and conduct a quantitative analysis to evaluate the integration level of Moroccans and Ukrainians in the different contexts of the Italian society and its main determinants (par. 4). The final section presents some reflections on potential links between integration policies and outcomes (par. 5).

2. Trends and characteristics of Moroccans and Ukrainians in Italy

Both the observed immigrant communities have significantly increased during the last 12 years (Figure 1). Moroccans, who already in the 1990s were found in large numbers, are around 510 thousand in 2013. They more than double Ukrainians (225 thousand), who mostly arrived in the last decade and increased after the regularizations.

Minors - arrived through family reunification channels or born in Italy - represent an important quota among Moroccans, while their numbers are negligible among Ukrainians, that migrate in Italy before 2010 mainly for labor reasons.

Figure 1 – *Trends of adult and total Moroccans and Ukrainians holding a residence permit (RP). Italy, 31st December 2001-2012. Absolute values*



Source: data of the Ministry of Interior revised and provided by ISTAT.

In the last 2 years, the increase of regular Moroccans was due both to minors who arrived in Italy for family reunification (more than 19 thousand in the period 2010-12) and, above all, to children born in Italy by Moroccan parents (more than 12.4 thousand in 2011 and almost 11.9 thousand in 2012). Recently, also Ukrainians slightly increased because of minors who, however, are still an extremely small proportion of the entire population (Moroccan and Ukrainian minors are respectively 30% and 9% of their reference population).

The two groups present a contrasting picture according to sex: women represent 44% of Moroccans and almost 80% of Ukrainians (Table 1), with differences that are amplified at specific adult age groups. The prevalence of men among Moroccans is larger in the 35-59 age group, as well as the predominance of women among Ukrainians is accentuated at older ages. The mean age of Moroccans is lower than 30 years, with no significant difference by sex (the mean age of men and women is respectively lower than 31 and higher than 28 years). The Ukrainian mean age is higher than 42 years and largely differs by sex (31.5 for men and 45.2 for women). Such difference is due to the low presence of Ukrainian women aged less than 18 (6% of them in respect to 24% of the male counterpart).

Table 1 – *Demographic characteristics of Moroccans, Ukrainians and other Third Countries' nationals who hold a residence permit (RP) or are registered with the parental one . Italy, 31th December 2012. Percentages and mean values.*

Demographic characteristics	Moroccans	Ukrainians	Other Third Countries
% women	43.9	79.8	48.0
% by age groups			
- under 18	30.8	9.2	24.1
- 18-34	28.2	19.9	32.4
- 35-54	33.0	47.4	35.6
- 55 and over	8.0	23.5	7.8
Mean age of women	28.5	45.2	32.0
Mean age of men	30.7	31.5	31.0
Dependency ratio	44.9	11.2	30.4
Child-woman ratio	45.0	5.6	23.6
% by geographic division			
- North-West	41.7	27.7	36.7
- North-East	31.3	25.2	27.9
- Centre	14.1	20.8	24.8
- South	9.3	24.6	7.4
- Islands	3.5	1.7	3.1
% in metropolitan provinces [(a)]	23.6	37.4	38.1

Note: (a) The twelve metropolitan provinces are: those related to the nine areas defined by Law 142 (i.e. the provinces of Turin, Genoa, Milan, Venice, Bologna, Florence, Rome, Naples and Bari) and three adding provinces in the islands (Palermo, Catania and Cagliari).
Source: our calculations based on data from the Ministry of Interior revised and provided by ISTAT.

The peculiarities by marital status well represent the age structure and the different cultural and migratory models which characterize the two groups. The majority of Moroccan men are single (53%) and the largest part of Moroccan women are married (47%). Also Ukrainian men are predominantly single (53%), while more than 7 out of 10 Ukrainian women are married or separated, divorced and widows (Table 2).

Table 2 – *Percentages by marital status of Moroccan and Ukrainian usual resident population divided by gender. Italy, 8th October 2011.*

Marital Status	Moroccans		Ukrainians	
	Men	Women	Men	Women
Single	53.2	45.7	52.9	26.8
Married	43.9	46.6	41.5	36.7
Separated/divorced	2.2	4.2	4.6	23.1
Widow	0.7	3.5	1.0	13.4

Source: our calculations based on 2011 Population Census.

The territorial distribution of the two groups largely differs among Italian regions: Moroccans live mostly in Northern Italian regions (more than 70%), while a significant proportion of Ukrainians lives in Central and Southern regions (more than 45%) and in metropolitan provinces (Table 1).

In addition to demographic and migratory characteristics, Ukrainians and Moroccans present very dissimilar behaviors in terms of employment, union and family formation and migration plans. As follows, it will be interesting to assess whether these differences play a significant role on the level of integration achieved.

3. Data and methods

Official statistics refer to the resident or regular population and do not provide enough information about life conditions and integration levels of immigrants. To overcome these limitations, we use the survey data carried out by the ISMU Foundation between the end of 2008 and the beginning of 2009 (Cesareo, Blangiardo, 2009). It includes 12 thousand adult immigrants living in 32 geographical units of the Italian territory (resident and non-resident, regular and irregular) and representative of the five different Italian geographical divisions. According to the un-weighted cases, interviewed Moroccans are almost 1,400 and Ukrainians almost 800. Collected information allow to conduct a detailed and multidimensional study of integration level of immigrants, overcoming the existing limits of available official data.

We consider 40 variables to define four composite indicators linked to four dimensions of integration: a cultural dimension, related to the language knowledge/use, the access to Italian news, the interest in the Italian events and the

sense of belonging to the Italian society; a social dimension, related to friendship relations, participation to group-associations, level of appreciation of the Italian lifestyle; a legal dimension, related to the legal status and the opinion about the importance to acquire Italian citizenship for themselves and their children; an economic dimension, related to the occupation, housing condition, saving capacities.

The modalities of each variable have been ordered according to an increasing level of integration. For each variable, we assign to each individual the higher score the larger is the quota of people who live in a worst condition of integration or, rather, the lower score the more numerous are those in a equal or better condition of integration. All variables' scores have been summarized by an arithmetic mean within each of the four observed dimensions, in order to estimate the relative indexes of integration, namely cultural integration, social integration, legal integration, economic integration. The values of indexes have been normalized between 0 and 1, that correspond to absence and maximum level of integration, respectively (for a more detailed description of the method see Cesareo, Blangiardo, 2009). The estimated indexes assume relative values that are comparable among sub-samples of interviewees according to their characteristics (e.g. citizenship, place of residence, education, occupation, etc.). Nevertheless, some data limitations persist: there is no way to consider autochthonous people and to conduct a longitudinal analysis.

4. The integration of Moroccans and Ukrainians: a comparative analysis

In table 3 we show the ranks of Moroccans and Ukrainians according to the mean scores obtained for the four dimensions of integration by the 17 most numerous national groups in Italy (Table 3). Generally speaking, Moroccans and Ukrainians rank very differently according to dimensions.

The Moroccan community is located in an intermediate position on the list. The worst performance is observed in the economic integration (15^{th} rank). Similarly, the mean cultural score (0.461) is lower than the average of immigrants (0.490).

Ukrainians lay close to the bottom of the rankings of all four dimensions of integration. The best performance is achieved in the cultural dimension although the score is only slightly higher than the national average (0.493). Living and working conditions are generally poor for such collective (economic integration), there isn't a significant participation in the social life of the country (social integration) and a significant interest to acquire Italian citizenship (legal integration). Adding elements come from multivariate analysis synthetically described below. Linear regression models consider adding predictors of the four indicators of integration separately for Moroccans and Ukrainians (see Table 4).

Table 3 – *Scores and ranks of Moroccans and Ukrainians in the four dimensions of integration. Italy, 2008-2009.*

	Integration dimensions			
	Cultural	Social	Legal	Economic
Scores of Moroccans	0.461	0.480	0.509	0.480
Scores of Ukrainians	0.493	0.437	0.406	0.503
Ranks of Moroccans among 17 groups[(a)]	12	5	4	15
Ranks of Ukrainians among 17 groups[(a)]	9	16	17	13
Minimum range scores	0.285	0.396	0.406	0.449
Maximum range scores	0.561	0.519	0.515	0.666
Mean observed scores	0.490	0.478	0.489	0.526

Note: [(a)] The seventeen immigrant groups come from: Albania, Bangladesh, China, Ecuador, Egypt, India, Macedonia, Moldova, Morocco, Peru, Philippines, Poland, Romania, Senegal, Sri-Lanka, Tunisia, Ukraine.
Source: our calculations based on ISMU data.

Table 4 – *Sign and level of significance of determinants of cultural, social, legal and economic integration of Moroccans and Ukrainians, according to linear regression analysis.*

Variables	Cultural integration		Social integration		Legal integration		Economic integration	
	Morocco	Ukraine	Morocco	Ukraine	Morocco	Ukraine	Morocco	Ukraine
Age	---	-		---		---	---	---
Lenght of stay	+++	+++	+++	+++	+++	+++	+++	+++
Lenght of stay (squared)	---		---		---		---	
Gender (ref = Men)								
- Women	+++	+++	+++	+	+++	+++		
Division of residence (ref = North)								
- Centre		++	+++					
- South		+++	++	+++		++	---	---
Municipalities (ref = Large)								
- Middle			+	+++	+++	+	+++	
- Small	--	---	-		+++		+	
Education (ref = Low)								
- Middle	+++		+++		+++		+++	++
- High	+++	+++	+++		++	-	+++	+++
Type of family (ref = Complete)								
- Single		-		---	---	---	---	---
- Divided	---	---	---	---	---	---	---	---
Remittances (ref = Regularly)								
- Occasionally	+					++	+++	+++
- Never	+				+++		+++	
Sence of belonging to the OC (ref = Much)								
- Somewhat	++	+	++	++				++
- Few	+++	++	+++	+++	+++			
- Nothing	++							
Interested in the OC (ref = Much)								
- Somewhat			++	+++	++	++		++
- Few		+++	++	++				+++
- Nothing	++	+++	+++	++				
R squared	0.295	0.199	0.191	0.263	0.362	0.251	0.233	0.183

Length of stay and age at migration are entered as continuous variables measured in years. We observe the persisting, positive and significant coefficients of the length of stay in all performed regressions. Conversely, the age of respondents assumes significant and slight negative coefficients in the four aspects of integration with few exceptions (it does not affect social and legal integration of Moroccans). According to such results, the integration of immigrants at very young ages is indispensable to reach the best performances. While women are favored in the first three dimensions of integration, there is no significant difference between sexes in economic integration. Living in the Southern regions is negatively associated with economic integration while, interestingly, has a positive effect in all other dimensions for both nationalities. In addition, integration is better in the urban centers of medium size. Living in small towns - probably characterized by higher "social control" - negatively affects the cultural integration of immigrants. Only for Ukrainians, education does not seem to play such a prominent role in integration processes, while the higher the level of education the lower the legal integration.

5. Concluding remarks on the link between integration outcomes and policies

Our findings show that the place of residence plays a fundamental role in determining integration outcomes regardless of integration dimension. Consequently, the role assumed (*de facto* and *de jure*) by local authorities in planning and implementing integration policies seems coherent and fully justified.

Looking specifically at Ukrainians and Moroccans' outcomes, our findings reveal as integration policies should first target their disadvantaged economic conditions. In the case of Ukrainians, there is a huge need of supporting women labour market trajectories, especially in Southern regions and large metropolitan areas. So far, integration policies towards women migrants conducted at a state level have, instead, mostly been directed towards improving linguistic and cultural integration with a specific focus on those women arrived for family reunification. On the contrary, it should be noted as Ukrainian women migrants' economic integration has been constantly supported through a variety of tools and initiatives designed at a local level. As with Moroccans, their difficult integration in the labour market deserves attention.

Morocco is, indeed, one of the few countries having signed a bilateral agreement on labour migration with the Italian government (year 2005). Among other objectives[2], such agreement provided migrants with specific tools (see e.g. pre-departure work and linguistic trainings) which – implicitly – would have supported their economic integration. Today, after 9 years, we may conclude that such instruments need, undoubtedly, to be revised and strengthened from an integration perspective. Cultural integration is another obstacle to Moroccan integration. However, while Italian policies

[2] For instance, the agreement aims at supporting seasonal labour migration and controlling irregular migration.

explicitly target the linguistic and cultural integration of reunified women – status which applies to the majority of Moroccan women –, our findings show that much more efforts should be instead put on supporting cultural insertion trajectories of the Moroccan male population. Ukrainians are found to be in a difficult position with respect to social and legal integration, too. The former can be partially attributed to the work conditions of a part of them (caregivers who cohabit with their employers and are often deprived of their autonomy and independence), while the latter is likely to be attributable to the fact that massive migration from Ukraine is a quite recent phenomenon. As a matter of fact, their relative position compared to other migrant (oldest) communities tends being weaker. It is worth noting, however, as both dimensions are almost absent from the integration political agenda at a state level.

References

BONIFAZI C., STROZZA S. 2003. Introduction, *Studi Emigrazione*, vol. XL, No. 152, pp. 690-696.

CESAREO V., BLANGIARDO G.C. (Eds.). 2009. *Indici di integrazione. Un'indagine empirica sulla realtà migratoria italiana*. Milano: Franco Angeli.

DALLA ZUANNA G., FARINA P., STROZZA S. 2009. *Nuovi italiani. I giovani immigrati cambieranno il nostro paese?*. Bologna: Il Mulino.

GOLINI A. (Ed.). 2006. *L'immigrazione straniera: indicatori e misure di integrazione*. Bologna: Il Mulino.

ZINCONE G. 2000. Introduzione e sintesi. Un modello di integrazione ragionevole, in G. Zincone G. (Ed.), *Primo rapporto sull'integrazione degli immigrati in Italia*. Bologna: Il Mulino, pp. 13-120.

SUMMARY

The present contribute aims to investigate the integration of Moroccans and Ukrainians in Italy according to a multidimensional approach. After briefly describing the trends and the demographic characteristics of the two communities, we use a multivariate approach to analyze the determinants of four dimensions of integration (cultural, social, legal and economic), in order to draw interesting conclusions in terms of policy recommendations.

Anna DI BARTOLOMEO, Migration Policy Centre, Florence, anna.dibartolomeo@eui.eu

Giuseppe GABRIELLI, University of Naples Federico II, giuseppe.gabrielli@unina.it

Salvatore STROZZA, University of Naples Federico II, strozza@unina.it

Rivista Italiana di Economia Demografia e Statistica Volume LXVIII n. 3/4 Luglio-Dicembre 2014

THE SECONDARY EDUCATION CHOICES OF IMMIGRANTS AND NON-IMMIGRANTS IN ITALY[1]

Michele Lalla, Elena Pirani

1. Introduction

Secondary schooling is not compulsory in the Italian educational system and enrolment decisions may originate from differences in individual behaviour or socio-economic conditions of families. Additionally, such decisions may affect opportunities for future employment and social mobility. All these aspects may differ among immigrant and non-immigrant youth and, for the former, secondary schooling plays a role in social integration as well (Entwisle and Alexander, 1993).

The objective of this paper is to ascertain the differences between the two groups, immigrants and non-immigrants (hereinafter referred to as Italians), with respect to the choice to continue or to interrupt their secondary schooling, taking into account individual, social and demographic characteristics and family background. The data were extracted from two surveys carried out by the Italian National Institute of Statistics (Istat): The European Union Statistics on Income and Living Conditions (EU-SILC) – carried out yearly since 2004 under the coordination of Eurostat – and the Italian Survey on Income and Living Conditions of the Families with Immigrants (IT-SILCFI) – carried out in 2009.

The paper is organised as follows. Section 2 concisely describes the theoretical background, and Section 3 illustrates the sample, data and some descriptive results concerning the main variables used in the subsequent analysis. Section 4 describes the models and includes comments on the results. Finally, Section 5 briefly concludes with some comments and remarks.

[1] This paper is based on data from Istat, European Union - Statistics on Income and Living Conditions (EU-SILC). The responsibility for all conclusions drawn from the data lies entirely with the authors: Disclaimer clause suggested by Eurostat on its website. The paper is the result of the cooperation of both authors. However, the specific contributions of each author are as follows: Sections 1 and 3 are by Michele Lalla, Section 2, 4, and 5 are by Elena Pirani.

2. Background

Educational choices of youths are made at a particular stage in the life of these young people, when influences both inside and outside the home are strongly felt. In this sense, educational choices strongly depend on both individual and family characteristics, as well as on the social and contextual background of the area where they reside.

Firstly, at individual level, gender, age and health conditions have proved to be associated with the choice to continue one's education and training. In this context, immigrant status, nationality, and the length of stay in the country clearly also play a role.

Secondly, educational choices reflect and originate from the family context of young people, including both natives and immigrants. The effect of family background on assimilation and expectations has been thoroughly analysed for both natives and immigrants, and different factors have been identified as relevant in these processes: household size and family composition, educational level of parents, socioeconomic status, parental language and expectations, parental supports and involvement, cultural background and income. The influence of these factors in the educational choices of young people has also been investigated (Luciano et al., 2009) to some extent.

Lastly, the social context of the community and the area of residence may be also relevant. The schooling context has been analysed as a source of inequality between natives and immigrants and/or among different groups of immigrants as well: attending kindergarten, previous experiences of success and failures, advice of teachers and peers, and availability of schools in the area. The context of the community of residence may refer to social characteristics of the neighbourhood (Pong and Hao, 2007) and to economic characteristics. The former have been often represented considering crime levels, characteristics of peers, companionship and so on, while the economic factors may refer to the employment/unemployment rate in the area of residence, the local gross domestic product, the value added by sectors (Bertolini et al., 2013).

3. Selected sample and variables

Our empirical analysis is basically based on the EU-SILC data gathered for Italy by Istat. The EU-SILC data refer to yearly information on nationally representative random samples of private households in each European country, and they comprise a cross-sectional and a longitudinal component (Eurostat, 2009).

Beside individual socio-demographic characteristics, EU-SILC provides micro-level data on income, poverty, social exclusion and living conditions. It started in 2004 under framework regulation (European Council) no. 1177/2003 adopted by the Council and the European Parliament in 2003.

In order to obtain a consistent sample and comparable information for immigrants, data from the Italian Survey on Income and Living Conditions of Families with Immigrants (IT-SILCFI) were considered together with the EU-SILC sample. This survey has the same structure as the EU-SILC survey, although it involves some additional specific variables. The IT-SILCFI was carried out by Istat only in 2009, so we decided to utilize data from both surveys for 2009.

Both surveys collected data at the household and the individual level. In 2009, the number of household units was 20,492 for EU-SILC and 6,014 for IT-SILCFI, while the number of eligible household members, i.e., people aged 16 and over, was 51,196 for EU-SILC and 15,036 for IT-SILCFI, for a total of 66,232 individuals (Table 1).

Overall, secondary education mostly involved youths under 20 years of age (Table 1). Only 2.6% (52 subjects) of those continuing their secondary education were 20 years old and 12.8% were aged 21 and over. The latter percentage was not too low, and it should be noted that 68 (26.5%) out of 257 subjects came from IT-SILCFI, and they were distributed over an age range of 21-61 years. There were 2,086 youths in secondary schools out of 2,675 in the sample, i.e. 78%.

Table 1 – *Number of subjects by type of school currently attended and age.*
Legend: SE=Secondary Education, TE= Tertiary Education, PhD= Philosophiae Doctor

ISCED Level	Age							Total	Sample
currently attended	<= 15	16	17	18	19	20	>= 21		16-19
Primary Education (PE)	0	**6**	**7**	**12**	**2**	4	112	143	**0**
Lower SE: 2-3Y (LSE)	0	**62**	**31**	**22**	**10**	6	62	193	**125**
Upper SE: 4-5Y (USE)	0	**546**	**488**	**467**	**193**	52	257	2,003	**1,694**
Post SE (no TE)	0	**3**	**1**	**3**	**10**	9	63	89	**17**
First/second-stage TE	0	**9**	**7**	**15**	**217**	253	1,658	2,159	**248**
Post TE	0	**1**	**1**	**0**	**0**	0	363	365	**2**
PhD	0	**0**	**0**	**0**	**0**	0	65	65	**0**
Not In School (NIS)	10,985	**69**	**111**	**141**	**268**	352	49,289	61,215	**589**
Total	10,985	696	646	660	700	676	51,869	66,232	**2,675**

The descriptive statistics for the main variables included in the analysis are reported in Table 2 and subdivided into three categories: (1) the socio-demographic characteristics of youths were gender, age, general health classified as in good versus not in good health, chronic illness and immigrant status both classified as yes or no; (2) parental and family information consisted of (for both mother and father) age, education (low, average, and high), and general health, employment

situation (either both or only one parent employed), and household income per capita; (3) lastly, due to the scarcity of detailed information, the geographic area of residence was simply defined through the degree of urbanisation (high, average, or low density) and the macro-region of residence (North-West, North-East, Centre, South, Islands).

Table 2 – *Mean of the main variables by status and by current education.*
Legend: LSS = Lower Secondary School, USS = Upper Secondary School, PSS = Post-Secondary School, NIS = Not In School. F = Father, M = Mother. Mdn=median.

Variables	**Non-immigrants = 73.5%**				**Immigrants = 26.5%**				
Current education	**LSS***	**USS**	**PSS**	**NIS**	**LSS**	**USS**	**PSS**	**NIS**	**Total**
No. of cases	64	1,293	240	370	61	401	27	219	2,675
Individual characteristics									
Women	0.47	0.51	0.54	0.45	0.46	0.51	0.67	0.46	0.50
Age	16.70	17.18	18.75	18.05	16.98	17.19	18.44	18.00	17.51
General health	0.06	0.04	0.06	0.08	0.02	0.03	0.07	0.05	0.05
Chronic illness	0.17	0.06	0.08	0.06	0.02	0.02	0.00	0.02	0.05
Parental and family characteristics									
Father's age	49.12	50.33	51.17	49.47	45.95	46.97	49.32	44.11	49.14
Mother's age	46.11	47.07	48.62	46.73	42.08	42.82	47.81	40.58	45.86
Max education (Mdn)	4	6	6	4	4	5	5	4	5
General health (Mdn)	4	4	4	4	4	4	4	4	4
Chronic illness	0.27	0.20	0.24	0.24	0.21	0.17	0.19	0.09	0.19
Employed: F & M	0.27	0.42	0.44	0.25	0.26	0.33	0.26	0.19	0.36
Employed: F	0.44	0.35	0.32	0.40	0.38	0.36	0.52	0.47	0.37
Employed: M	0.13	0.13	0.12	0.13	0.20	0.20	0.07	0.22	0.15
Retired	0.05	0.04	0.05	0.08	0.00	0.00	0.00	0.00	0.04
Other condition	0.13	0.07	0.07	0.15	0.16	0.10	0.15	0.12	0.09
Permanent job	0.58	0.65	0.68	0.53	0.16	0.22	0.19	0.21	0.52
White-collar	0.27	0.46	0.49	0.22	0.05	0.10	0.33	0.03	0.33
Income per capita/10^3	9.30	10.73	12.26	8.50	4.50	6.56	9.57	5.63	9.33
Area of residence									
City: High density	0.39	0.34	0.35	0.35	0.38	0.39	0.48	0.38	0.36
City: Average density	0.31	0.40	0.45	0.36	0.46	0.46	0.41	0.43	0.41
City: Low density	0.30	0.26	0.19	0.28	0.16	0.14	0.11	0.18	0.23
Region: North-West	0.22	0.18	0.18	0.17	0.16	0.24	0.33	0.21	0.19
Region: North-East	0.41	0.22	0.18	0.18	0.20	0.25	0.22	0.21	0.22
Region: Centre	0.17	0.21	0.27	0.17	0.20	0.24	0.15	0.15	0.21
Region: South	0.16	0.28	0.30	0.33	0.26	0.15	0.22	0.29	0.27
Region: Islands	0.05	0.11	0.08	0.14	0.18	0.12	0.07	0.14	0.11

In general, women tended to continue their education longer than men ($\chi^2_3 =$ 8.60 with $p<0.035$). Women attending upper secondary or post-secondary education represented 64.4% and 11.1% of the sample, respectively, with respect to

62.2% and 8.9% registered for men. The percentage of women not in school was lower than that of men: 20.2% versus 23.9%.

Young immigrants tended to continue their education less than young Italians do (χ^2_3 = 110.27 with p<0.000). Only 3.3% of immigrants attended lower secondary education in 2009, with respect to 8.6% of Italians; the percentages of immigrants attending upper secondary or post-secondary education were lower than those of Italians: 56.6% versus 65.7% and 3.8% versus 12.2%, respectively. On the other hand, the percentage of immigrants not in school was disproportionately higher than that of Italian young people (30.9% versus 18.8%).

The general health of youth was weakly associated with their enrolment in school. Youths in bad health tended to prolong their education less than those without health problems. The presence of chronic illnesses did not appear to be associated with educational decisions in a relevant manner.

Differences between young Italian natives and immigrants were also found for parental background. The age of fathers and mothers of Italians was significantly higher than that of fathers and mothers of immigrants, showing on average a difference equal to 4.1 years ($F_{7;2667}$= 36.78 with p<0.000) and 4.9 years ($F_{7;2667}$= 45.19 with p<0.000), respectively. Italian parents seemed to be affected by chronic illness more than immigrant parents. The economic status and the occupational status of Italian fathers and mothers were significantly higher than that of immigrant parents, as was the total family income per capita of Italians: 4,244€ ($F_{7;2667}$= 34.61 with p<0.000). Moreover, this income is almost halved for youths attending post-secondary education and for youths who were not enrolled in schools.

4. Results

The decision to continue or interrupt one's education was analysed for young Italians and immigrants. A binary variable, *Y*, denoting the dichotomised choice with respect to schooling, "in school" (*y*=1) versus "not in school" (*y*=0) was considered with respect to a vector of covariates **X**. Let $\pi(\mathbf{x})$ be the probability that *Y*=1 depending on the vector of covariate values **x**. The logit model is

$$\pi(\mathbf{x}) = \frac{\exp(\mathbf{x}'\boldsymbol{\beta})}{1+\exp(\mathbf{x}'\boldsymbol{\beta})} = \Lambda(\mathbf{x}'\boldsymbol{\beta}) \quad (1)$$

where $\Lambda(\cdot)$ denotes the logistic cumulative distribution function and the vector of coefficients $\boldsymbol{\beta}$ describes the effect of the covariates **X** on $\pi(\mathbf{x})$.

The covariates were selected based on the literature and depending on their statistical significance. Our main explanatory variable was the immigrant status of young respondents. Given that our objective was to ascertain differences between immigrants and Italians with respect to the choice to continue or to interrupt their education, we sequentially estimated models with different sets of covariates. Among the individual socio-demographic variables, the following were selected: immigrant status, gender and age, while personal health or any chronic illness or unmet need for medical and dental examination or treatment were excluded to avoid capturing effects concerning relatively few individuals. The geographic context was introduced, differentiating among the Italian macro-regions (North, Centre, South). Family background was introduced in the logit models through the variables concerning mother's and father's age, their educational level, activity status (differentiated by occupation and type of occupation) and self-perceived health. The logarithm of total income was introduced to account for the economic situation of the household. Considering the variables included in the model, the reference individual was an Italian male, living in the North of Italy, having parents with average schooling, both employed, none of them white-collar, and in good health. Table 3 reports the odds ratios (OR) and p-values (Pr>z) of the estimated models.

The young immigrants revealed a significant lower probability of continuing their education than young Italians: controlling only for gender, age and macro-area of residence (Model 1). Young immigrants were at risk of not remaining in school, i.e., 50% less than their Italian counterparts. However, the magnitude of the effect – and its significance – slowly decreased as other covariates were taken into account. In the completed model, which controlled for all parental and family covariates (Model 4), the probability of young Italians continuing their education was not significantly higher than that of young immigrants. The variables used to represent the family environment appeared to play a relevant role in explaining differences in continuing education between natives and immigrants. Firstly, for youths with parents having a high (low) level of education, the probability of continuing their education was twice (half), OR=2.03 (OR=0.52) that of youths with parents having average schooling. Secondly, with respect to individuals whose parents were both employed, the probability of continuing education was lower for other parental employment situations (with ORs ranging from 0.48 to 0.73). Thirdly, the type of occupation also mattered: having at least one parent employed as a white-collar worker increased the probability of being enrolled in school, with respect to less skilled occupations. Finally, having parents in good health represented a further factor that might contribute to enhancing school enrolment.

Table 3 – *Estimated odds ratio (OR) and corresponding p-values (Pr>z) for some models.*

	Model 1		**Model 2**		**Model 3**		**Model 4**	
	OR	**Pr>z**	**OR**	**Pr>z**	**OR**	**Pr>z**	**OR**	**Pr>z**
Immigrant: yes=1, no=0	**0.48**	**0.000**	**0.47**	**0.000**	**0.69**	**0.002**	**0.90**	**0.439**
Women: yes=1, no=0	1.22	0.046	1.21	0.049	1.21	0.065	1.28	0.029
Age of respondent	0.56	0.000	5.27	0.353	7.58	0.279	12.73	0.212
Age of resp. (squared term)			0.94	0.211	0.93	0.154	0.91	0.114
Region: Centre			1.18	0.248	1.26	0.127	1.34	0.074
Region: South & Islands			0.66	0.000	0.82	0.102	0.91	0.475
Age father					1.09	0.065	1.12	0.117
Age father (squared term)					1.00	0.306	1.00	0.276
Age mother					1.11	0.005	1.11	0.050
Age mother (squared term)					1.00	0.017	1.00	0.114
P education: Low					0.39	0.000	0.52	0.000
P education: High					1.94	0.002	2.03	0.003
P occupation: Only father					0.69	0.004	0.73	0.023
P occupation: Only mother					0.52	0.000	0.56	0.008
P occupation: Both "other"					0.48	0.000	0.48	0.001
P Typ_Occ: White-collar							1.81	0.000
P Health: Less than good							0.76	0.025
Income (log)							1.03	0.676
Constant	100.00	0.000	0.00	0.614	0.00	0.332	0.00	0.229
Pseudo R^2 – no. of cases	*0.078*	*2,675*	*0.087*	*2,675*	*0.161*	*2,307*	*0.170*	*2,307*

Legend: P = Parents, Typ_Occ = Type of Occupation, "other"=other status.

5. Conclusions

An empirical analysis was performed to investigate differences in educational enrolment between native Italian young people and immigrant youth. Our empirical results are coherent with those previously reported in the literature, and suggest that an "immigration" gradient is present in educational choices also in Italy. However, differences among the two groups disappear when family background is taken into account. Most of the differences in educational enrolment between Italians and immigrants were absorbed by the socio-economic status of their parents, i.e., their level of education, employment status and occupational position. These results highlight the need for integrated policies in educational programs, directed both at sustaining youth and helping their families, in order to enhance and improve educational enrolment of young immigrants and foster a complete integration process. However, further investigation is needed to analyse potential differences more thoroughly at the geographic level.

References

BERTOLINI P., LALLA M., PAGLIACCI F. 2013. School enrolment of first- and second-generation immigrant students in Italy: A geographical analysis, *Papers in Regional Science*. (DOI: 10.1111/pirs.12042).

ENTWISLE D. R., ALEXANDER K. L. 1993. Entry into school: The beginning school transition and educational stratification in the United States. *Annual Review of Sociology*, Vol. 19, pp. 401-423.

EUROSTAT 2009. *Description of Target Variables: Cross-section and Longitudinal*, EU-SILC 065 (2009 operation). Directorate F, Unit F-3.

LUCIANO A., DEMARTINI M., RICUCCI R. 2009. L'istruzione dopo la scuola dell'obbligo. Quali percorsi per gli alunni stranieri? In ZINCONE G (ed.) *Immigrazione: segnali di integrazione. Sanità, scuola e casa*. Bologna: il Mulino, pp. 113-156.

PONG S., HAO L. 2007. Neighbourhood and School Factors in the School Performance of Immigrants' Children. *International Migration Review*, Vol. 41, 1, pp. 206-241.

SUMMARY

The secondary education choices of immigrants and non-immigrants in Italy

The choice of secondary schooling, which is not compulsory in Italy, is important for youths because it affects future opportunities for employment and social mobility. Secondary schooling also plays a role in the social integration of immigrants. To ascertain the presence of differences between young Italian natives and immigrants in education choices, two datasets for 2009 were used: the European Union Statistics on Income and Living Conditions (EU-SILC) and the Italian Survey on Income and Living Conditions of the Families with Immigrants in Italy (IT-SILCFI).

Analysing a sub-sample of young Italians and immigrants, aged between 16 and 19 years old, the association of both individual and family explanatory variables with the choice of secondary schooling (yes/no) was assessed using logistic models. The results show that young immigrants tend to interrupt their schooling earlier than their Italian counterparts. However differences disappear when family background and parental characteristics are taken into account.

Michele LALLA, Full Professor of Social Statistics, University of Modena and Reggio Emilia, michele.lalla@unimore.it

Elena PIRANI, Assistant Professor of Demography, University of Florence, pirani@disia.unifi.it

INDIVIDUAL AND HOUSEHOLD CHARACTERISTICS AND MIGRATORY MODELS OF IMMIGRANTS IN CAMPANIA[1]

Alessio Buonomo, Elena de Filippo, Giuseppe Gabrielli

1. Introduction

Campania, together with some other South regions in Italy, was characterized at the beginning of the phenomenon by a temporary foreign presence. Subsequently, we observed a scenario of an increasing stability, if we consider specific immigrant groups, together with the arrival of immigrants coming from new areas of origin (Ammaturo, de Filippo, Strozza, 2010).

The data sources on the presence of foreigners in Italy appear nowadays more and more rich, articulate and reliable than in the past. The official data contribute to outline the main demographic characteristics of the foreign population resident in Italy and its territorial divisions.

The immigrant residents in Campania have increased in the period 2001-2013 from more than 40thousand to more than 170thousand; today they represent 4% of the Italian foreign presence and 3% of the regional population. The immigrants reside mainly in the province of Naples (48%), but also significantly in the provinces of Salerno (23%) and Caserta (19%). The age structure of the foreign population resident in Campania show a gender imbalance in favor of women (57% of women). The average age of men and women is respectively 33 and 37 years old. Generally speaking, the quota of immigrants coming from European Third Countries is higher (31%) than Africans (17%). Romania and Morocco are two of the three most representative citizenships among immigrants in Campania. The Albanians represent only the 4% of foreign presence, while Ukrainians are the largest group (22%).

We need to use *ad-hoc* sample surveys to consider the non-resident population (in particular the irregular one) and to investigate the living conditions, the family characteristics and the migratory models at local level (Strozza et al., 2002).

[1]This work is the result of a close collaboration between the authors. As for the actual text, paragraph 2 has been written by A. Buonomo; paragraph 4 by E. de Filippo; paragraphs 1 and 3 by G. Gabrielli.

The present paper aims to advance knowledge on the characteristics of foreign presence in Campania and their migratory models. Using a quantitative approach, two-way analyses are followed by a factorial analysis to provide a synthetic picture of the observed phenomena.

The data is originated by a recent sample survey about immigrants in Campania conducted by Dedalus, in collaboration with the University of Naples Federico II, and granted by the Regional Service of Cultural Mediation (Por Campania FSE 2007-2013). The survey was conducted, through 72 questions, between May and October 2013 and regards 3,816 adult immigrants (de Filippo, Strozza, 2014). Data is weighed using "center sampling technique" based on a double set of weights in order to make the results representative of the local migrant presence (Baio, Blangiardo, Blangiardo, 2011).

2. A synthetic and descriptive overview of immigrant scenario in Campania

In the sample survey of 2013 on the foreign presence in Campania (legal and no legal resident), the majority of respondents come from Central and Eastern European countries; anyway, a significant quota is represented by Africans (23%) and Asians (20%). In table 1, Ukrainians represent the most numerous group (22%) followed by Romanians (18%). Both the communities show a gender imbalance in favor of women (respectively 76% and 59%). Moroccans represent the third group by number of presence and male majority (25% are women). The 65% of adults are aged between 25 and 44 yrs. The highest elderly ages are among Ukrainians, Russians and Polish; while the youngest groups are Chinese, Bengalese and Moroccans. The majority of respondents (60%) has at least a graduation (43% diploma and 17% university degree). The less educated immigrants are Moroccans, Senegalese and Bengalese; while Russians and Ukrainians have the highest levels of education with university degrees (respectively 47% and 36%).

One third of respondents (28%) doesn't have a regular resident permits (Table 2). Russians and Senegalese assume the highest quota (respectively 36% and 37%), while Chinese the lowest one (13%). The 38% of immigrants has irregular employments or self-employments. Senegalese, Polish, Romanians and Albanians assume the highest quota of irregular jobs among the others. Considering the high quota of un-declared work, it is not surprising that only 77% of respondents express their income. Chinese have the largest quota, among respondents, with a mean household equivalent income higher than 1thousand euro (46%). Conversely, 70% of Senegalese have household incomes lower than 500 euro. The majority of immigrants lives in rented houses and often without a regular contract. Two thirds live in independent residences, thus living in accommodations owned or rented alone or with their family; among them, almost all

of the Chinese (93%) live in this condition. According to a deprivation index (see note in table 2) almost 35% of respondents have a medium level of deprivation while one foreigner out of five (20%) has a high level of deprivation. Nearly half of the Senegalese have a high level of deprivation while the Chinese have a null level.

Table 1 – *Top ten nationalities of the foreign population by selected individual characteristics. Campania 2013. Absolute values and percentages.*

Citizenships	N	% of women	Mean age at interview	% of graduates
Ukraine	841	76.1	42.6	86.0
Romania	675	59.1	38.0	51.1
Morocco	361	24.7	35.4	29.6
Sri-Lanka	218	42.9	39.1	62.1
Poland	217	73.7	40.9	74.8
China	202	44.1	33.3	62.6
Albania	129	37.2	38.8	53.7
Senegal	86	14.0	38.7	35.0
Russia	77	85.7	42.1	86.9
Bangladesh	72	11.1	33.2	49.5
Other	938	45.2	37.6	44.7
Total	3,816	53.1	38.7	58.2

Table 2 – *Top ten nationalities of the foreign population by selected socio-economic characteristics. Campania 2013. Absolute values and percentages.*

Citizenships	% of no legal residents	% of irregular workers	Mean household equivalent income (euro)	% of independent residence	% of medium or high deprivation level[(a)]
Ukraine	24.2	31.7	666.26	63.1	43.1
Romania	33.0	49.8	627.95	60.6	46.7
Morocco	28.3	38.0	549.38	73.0	66.1
Sri-Lanka	26.0	26.0	517.73	68.4	68.2
Poland	29.4	46.8	725.59	55.3	46.6
China	13.4	5.5	1043.15	92.7	26.0
Albania	21.2	49.7	672.52	89.0	64.2
Senegal	37.2	58.2	459.13	55.5	83.9
Russia	36.5	42.0	648.38	53.6	47.0
Bangladesh	27.5	42.1	533.86	50.8	53.6
Other	31.2	40.4	558.65	58.3	67.7
Total	28.2	38.3	627.55	64.1	54.6

Note: a) According to Istat (2011),we consider 9 "deprivations" or difficulties of everyday life. Low deprivation means one or two deprivations;medium deprivation mean between three and five deprivations; high deprivation means at least six deprivations.

The 35% of respondents have been present in Campania since a period of 5-9 years, the 27% since 10-14 years and almost the same quota is for recent arrivals (0-4 years). In table 3, the Albanians are on average the most ancient group (more than 12 years), while Russians and Romanians are those of the most recent arrival. The 64% of immigrants is forerunner (see note in table 3). Chinese present mainly a migratory family model (only 32% of them have no relatives on arrival); conversely, Bengalese present the highest quota of forerunner (80%) showing mainly an individual migratory model. More than 80% of immigrants don't have the intention of living in Italy within 1 year; such percentage is halved if we consider a period of 5 years (44%). The Sri-Lanka immigrant group assumes the highest quota of those who intend to remain in the region; while Senegalese and Bengalese present the lowest one.

In table 4, immigrants in Campania live mainly as a couple (35%) or with friends and/or other relatives (35%). Russians confirm an individual migratory model (53% express to live alone). The largest quota of Bengalese, Senegalese and Moroccans live with friends and/or other relatives. The migratory model of Chinese and Albanians is strongly family centered (more than 50% of them live as couples). The 62% of immigrants have at least one child, but only 30% have children born in Italy. Ukrainians assume the highest quota of respondents with at least one child (76%), but the second lowest percentage of respondents with children born in Italy. Interestingly, among Chinese, less than 60% has at least one child (the fourth lowest value), and the 57% have children born in Italy (the second highest quota).

Table 3 – *Top ten nationalities of the foreign population by selected migratory characteristics. Campania 2013. Absolute values and percentages.*

Citizenships	Length of stay(yrs)	% of forerunner[(a)]	Intention of living in Italy within 12months (%)	Intention of living in Italy within 5 years(%)
Ukraine	8.9	68.4	85.5	47.2
Romania	6.6	63.6	84.0	47.0
Morocco	9.0	60.6	76.5	41.3
Sri-Lanka	9.2	44.6	93.2	58.6
Poland	9.3	73.5	87.3	39.8
China	7.7	32.1	84.7	50.8
Albania	12.6	43.2	84.0	56.5
Senegal	10.2	75.1	67.0	36.0
Russia	6.5	75.2	77.9	45.1
Bangladesh	8.1	80.1	73.3	39.2
Other	8.6	69.8	73.0	37.0
Total	8.5	63.8	80.9	44.3

Note: a) We consider as forerunner the migrant without any family member at arrival already settled in the area.

Tabella 4 – *Top ten nationalities of the foreign population by selected household characteristics. Campania 2013. Percentages.*

Citizenships	Household (row %)			% of respondents with at least one child	% of respondents with at least one child born in Italy
	Single	Couple	Other		
Ukraine	43.6	33.5	22.9	75.6	17.8
Romania	37.7	36.3	26.0	62.4	25.5
Morocco	17.5	29.0	53.5	43.6	43.4
Sri-Lanka	13.6	44.6	41.8	56.7	34.4
Poland	36.1	31.5	32.5	63.7	32.4
China	11.7	58.8	29.5	59.3	56.8
Albania	9.9	62.1	28.0	65.8	61.7
Senegal	9.3	23.8	66.9	61.6	25.0
Russia	52.7	17.4	29.9	65.1	1.4
Bangladesh	6.4	24.2	69.4	53.7	33.4
Other	28.6	32.3	39.1	58.2	36.1
Total	30.1	35.4	34.5	62.1	30.4

3. A synthetic picture: the categorical principal component analysis

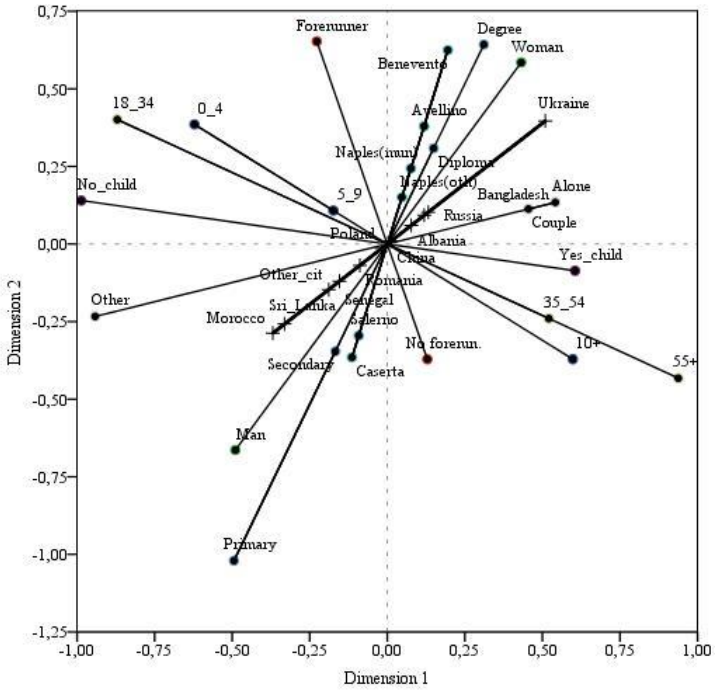
In order to synthesize the main significant aspects reported above, we conducted a factorial analysis namely: categorical principal component analysis. Generally speaking, the purpose of such analysis is to reduce the number of dimensions, thus the number of observed characteristics. In other words, such statistical method allows to identify the dependence structure among modalities of categorical and qualitative variables through their graphical representation on a space of minimum dimensionality. The axes of this bi-dimensional space are the result of the best synthesis of the information referred to each modality of considered variables. Excluding the socio-economic characteristics of respondents, we include in the analysis 1 illustrative variable (country of citizenship) and 8 active ones (see note in figure 1)[2]. The two observed dimensions of the factorial plane explain the 43% of the total variance (Table 5). The first dimension, that explains the 27% of inertia, is linked mainly to time and household characteristics. This dimension is therefore primarily characterized by the aspects related to the evolution of the person-time or to the life course of individuals and their families. The second dimension, that represents the 16% of inertia, is described mainly by individual characteristics and migratory models. These characteristics define situations almost always invariant over time.

[2]The "active" variables contribute to define the observed dimensions; the "illustrative" variables don't define the observed dimensions, but allow to better explain the meaning of them.

Table 5 – *Categorical principal component analysis. Total variance and explained inertia by active variable and dimension. Percentages.*

Characteristics	Variables	Dimension 1	Dimension 2
Time character.	Length of stay	11.7	7.8
	Age at interview	24.1	8.9
Household character.	Type of Household	21.7	2.3
	At least one child *vs.* no child	27.8	1.0
Individual character.	Women *vs.* men	9.8	31.1
	Education	3.1	22.7
Migratory character.	Province of residence	0.4	6.7
	Forerunner *vs.* no forerunner	1.4	19.5
Total		100.0	100.0
% of explained variance		*26.9*	*15.6*

Figure 1 – *Categorical principal component analysis. Factorial plane.*



Note: In the factorial plane we include 1 illustrative variable (country of citizenship) and 8 active ones namely: length of stay in Italy (0_4, 5_9 and 10+); age at interview (18_34, 35_54 and 55+); type of household (alone, couple and other); to have at least one child (yes_child, no_child); gender (women, men); education (primary, secondary, diploma and degree); province of residence (Benevento, Avellino, Caserta, Salerno, Naples municipality and the rest of Naples' province); to be forerunner (forerunner, no forerun).

The top ten citizenships are reported on the factorial plane (Figure 1) according to a trajectory that places Moroccans and Ukrainians at both ends. The latter nationality seems to distance itself more than other observed groups. Some characteristics are placed along similar trajectories to the one of citizenship. They are, first of all, the categories related to gender and education: the modalities "woman", "diploma" and "university degree" are placed in the first quadrant, close to Ukraine; on the opposite, "man", "primary" and "secondary" school, together with household with friend and/or other relatives (category "other"), are placed in the third quadrant, close to Morocco. An adding variable is the province of residence that assume the same trajectory of the ones reported above. The analysis places Benevento, Avellino and Naples in the first quadrant; Caserta and Salerno in the third one. Analysis confirms the strong differences existing between Ukrainians and Moroccans, that assume very particular and specific characteristics. Between such dichotomy, we observe, however, a number of intermediate positions, defined on the one hand by the aspects examined so far and, on the other, by the other variables included in the factorial plane. We leave the reader to draw any further comments that we cannot report here for space reasons.

4. Conclusions

The above data are indicative of an "archipelago" of immigrant people, consisting not only of a significant variety of backgrounds, but also by different experiences and projects of foreign presence in Campania.The vocation of the region as area of transit and temporary stay is undoubtedly lower than that of the past (even if it has not disappeared completely), in favor of more permanent immigration.

Specific groups, more than others, continue to consider Campania as area of transit. Such behavior only in part is due to the phase of migration and to the length of stay of the specific group or of the individual. There is the case of Senegalese, for example, that arrived for the first time in Campania during the 80s' and have never, or just shortly, started a settlement process on the region. Generally speaking, they continue to live the temporary immigration in Campania, as a stage of a migration project oriented elsewhere or of a "circular" migration (investing economic and affective capitals in the country of origin). On the contrary, Chinese, a community of recent arrival, already assume several elements of the long and complex integration process (high percentages of legal resident, regular job, high income, presence of family households, children born in Italy, intention to live in Campania in the near future …). In other cases, as for the Sri-Lankan, the migration seniority corresponds to a lower propensity to leave the region and, in general, a greater stability.

The main determinants of migratory models continue to be, in addition to the characteristics of immigrants and of the related communities, the opportunities offered by the welfare system and the labour market. If in the past, in Campania, they were push and pull factors of migration due to their weakness, today - because of the economic crisis and the lack of opportunities elsewhere - they have become the context, not without contradictions, within which the process of stabilization and integration is emerging in the region.

Essential bibliographic references

AMMATURO N., DE FILIPPO E., STROZZA S. (a cura). 2010. *La vita degli immigrati a Napoli e nei paesi vesuviani: un'indagine empirica sull'integrazione,* Milano: Franco Angeli.

BAIO G., BLANGIARDO G.C., BLANGIARDO M. 2011. Centre sampling technique in foreign migration surveys: A methodological note. Journal of Official Statistic, 27, 451-465.

ISTAT. 2011. *I redditi delle famiglie con stranieri. Anni 2008-2009*. Statistiche Report, 22.

STROZZA S., NATALE M., TODISCO E., BALLACCI F. 2002. *La rilevazione delle migrazioni internazionali e la predisposizione di un sistema informativo sugli stranieri*. Rapporto di Ricerca della Commissione per la Garanzia dell'Informazione Statistica, 02.11.

DE FILIPPO E., STROZZA S. (a cura). 2014. *Indagine sulla presenza straniera e il livello di integrazione degli immigrati stranieri presenti nella regione Campa*nia, Rapporto di ricerca progetto Yalla, Napoli.

SUMMARY

Individual and household characteristics and migratory models of immigrants in Campania

The need to analyze migration patterns and define the socio-demographic characteristics of migrants and their families in different local contexts of Italian society has become, in the recent years, an important aspect in parallel with the increase of foreign presence and its gradual settlement in the destination area. Adopting a quantitative approach, this contribution aims to provide a summary report on individual and household characteristics of the foreign presence in Campania and migratory models.

Alessio BUONOMO, Università di Napoli Federico II, alessio.buonomo@alice.it
Elena DE FILIPPO, Cooperativa Dedalus, Napoli, presidenza@coopdedalus.it
Giuseppe GABRIELLI, Università di Napoli Federico II, giuseppe.gabrielli@unina.it

Rivista Italiana di Economia Demografia e Statistica Volume LXVIII n. 3/4 Luglio-Dicembre 2014

LOOKING BACK TO LOOK FORWARD: THE ITALIAN ACTIVE AGEING IN BETWEEN THE OLD AND THE NEW MILLENNIUM

Francesca De Palma, Stefania Girone, Sara Grubanov-Bošković

1. Introduction

According to the World Health Organization, active aging implies the optimization of physical, social and mental health opportunities which enable older people to play active roles in society as well as enjoy an independent and quality life. Based on that definition, during the last two decades, most of the developed countries have gradually consolidated strategies and planned means in order to push older people to carry on social and work activities (CE, 2008). Thus, different policies and practices on urban planning, rural development, access to health care, family, education, social security, employment, social engagement, free-time, and so on, have been reviewed in order to identify those suitable tools that allow older people to age in good health (enabling them to actively support both labor market and societies), as well as allow to cope with demographic challenges in a fair and sustainable way for all generations.

Although the Italian legislative framework in terms of active ageing is still quite scarce and rather fragmented at local level (Mirabile *et al.*, 2009; Ciccarone, 2012), we can shortly say that the main areas involved in its action range regard: 1) participation to the labour market; 2) lifelong learning; 3) active engagement; 4) health and quality of life; 5) transportation and mobility services.

Based on what has just been stated, we intend to assess how much those dimensions solely related to health, active social-participation and lifelong learning have influenced the choice of the Italian male population[1], aged between 55 and 75 years old, to remain active on the labor market over the past two decades, thus postponing transition to inactivity.

[1]The choice to only take into consideration the male population is due to the different Italian pension provision deserved to both sexes. In response to the European Commission requests, Italy has just recently (in 2010) made the first legislative step towards a gradual equalization of males' and females' pension provisions.

In view of planned policy-interventions and empirical research-outcomes, we assume that the worsening of health-conditions among older males will result in a greater propensity to anticipate inactivity-status (Zucchinelli *et al*., 2010; Innocenti, Vecchiato T., 2013). Moreover, by taking into account health improvements among the Italian population which, e.g., have entail significant increases in the residual-life-average among the 55 year-old people (in the case of the male population, it rose from 23.1 years in 1993 to 26.7 years in 2012 (Istat), we assume that there is a much greater capacity to lengthen the labour-cycle nowadays than there was in the past.

The "active social participation" issue, thus far considered as "social engagement", will be even analyzed by considering the different historical backgrounds and, therefore, the policies in force at that very moment. In this regard, based on some empirical researches (Attwood *et al*. 2003; Population Reference Bureau, 2011), we assume there is a positive relationship between active-status on the labor market and social engagement.

Finally, we will examine what role the lifelong learning (assessed by the use of modern technology, e.g., computer, internet, mobile phones, etc.) plays in the choice, made by the 55 year-old Italian males, of keeping up active. Previous researches have positively associated the use of technology to the prolongation of activeness on the labor market (Ala-Mutka *et al*., 2008; Peacock, 2009) and, in this sense, the 2012 statistical data enables us to appraise how much technological-means are crucial for the Italian older males in order to keep their active-status up.

2. Reference data

This analysis takes into account the database of the *Indagine Multiscopo sulle Famiglie* (Multipurpose Survey for Households) - carried on by the Italian Institute of Statistics - which collects important information about Italian families' everyday life. Furthermore, since Italian pension-provision differs by gender, for the purpose of this study only the male population has been taken into consideration: thus, the sample made up of 5,382 males refers back to the 1993 Survey, while the one of 5,570 males, to the last 2012 Survey[2]. Before examining the determinants that have pushed older males to keep up active on the labour market, it is necessary to specify what we exactly

[2]Yet, there are many other surveys focused on active aging such as *The Survey on Heath, Ageing and Retirement in Europe* (SHARE) which collects a large amount of data and allows comparisons among countries involved in this initiative. However, this particular survey provides information since 2004/2005 only (i.e., the first year that the survey was carried on) and, for this reason, does not fit our kind of analysis aimed at grasping the changes of the Italian population over twenty years, i.e., in the transition between the Twentieth century and the Twenty-first one.

mean by "older population" and "active status on the labour market". Generally, "older population" refers to the ones aged between 55 and 75 years old, as 55 represents the minimum age-threshold at which people can be categorized as "older workers" or "emerging older generation", and 75 is the maximum age-threshold which marks the end of the "older" condition and the beginning of the "very old" one[3]. Besides, the "active on labour market status" is generally given to those who declare in the survey to "be employed" or "be looking for a job", while the ones belonging to other categories (disable to work, retired, and other conditions) are considered "inactive". As explicative variables, we have considered three different dimensions: the first one is related to socio-territorial aspects (age, current residence, education, family); the second one regards healthcare and other features involved in political actions which strongly encourage active aging (health condition, social participation, economic status); the third one concerns lifelong learning and is assessed by the use of technology (mobile phones, PCs, iPads, etc.) which - particularly during the last two decades - have pushed lots of seniors to acquire skills to manage it. Yet, technological aspects are available just from the 2012 Survey.

3. The active aging in Italy: its determinants in the past and at present

As the descriptive analysis shows, afterward the Italian social security reforms (aimed at delaying the exit from productive lifecycle either by raising the retirement age and/or improving the health conditions) were implemented, the quota of those males aged 55-75 years old and active on labor market, has boosted from 30.7% in 1993 to 35.1% in 2012.

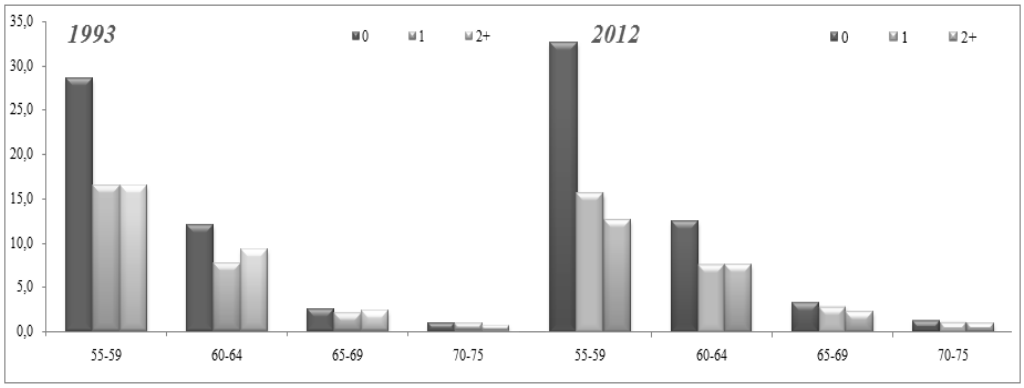
The profile of the "healthy older worker" clearly stands out in both years (Figure 1): it refers to a "young older person" aged 55-59 years old, who has not suffered from any chronic disease, and therefore has a positive view/opinion of his own health condition. Besides, the 2012-data shows that over one third of the whole male-sample uses Internet, half of which is active on the labor market, thus shaping the profile of a "web-surfing older worker".

As the analysis gets on with the binomial logistic regression, "age" clearly becomes the most important variable of the model. Thus, as age raises, the propensity to remain active on the labour market decreases. When age is divided into five-year groups and

[3]In a longitudinal perspective, this means that data will highlight active-condition for two different cohorts: the one born by the end of the World War I (1918-1938) and collected by the 1993 Survey; the other one born during the pre- and post-World War II period (1937-1957) and collected by the 2012 Survey.

the "55-59 age-threshold" is taking as reference, it comes out that the propensity to remain active drastically falls down at the next age-threshold (i.e., 60-64 years old), thus reaching very low values as it goes on at older age-groups: besides, this kind of relationship seems to be more pronounced in 2012 than in 1993.

Figure 1 – *Active older males by age and number of chronic diseases contracted.*



*Source: own elaborations based on the data set "Indagine Multiscopo sulle Famiglie", 1993 and 2012***.**

For what concerns healthcare, data confirms that the worsening of males' health conditions (both objective or subjective) lowers the propensity to remain in the workforce. Both years (1993-2012) show no changes with regard to the relationship between work and health conditions, measured in terms of presence of chronic diseases: e.g., the propensity to remain active is about 31.0% lower for those who suffered from a chronic disease compared to those who have never contracted any.

On the other hand, according to the perceptions males have about their own health conditions, outcomes reveal some important differences: in fact, in 2012, the ones who negatively judged their health-status disclose a 27.0% lower active-attitude than the ones who positively assessed it, whereas in 1993 this gap was much more marked.

Moreover, even the relationship between activeness on the labor market and social engagement seems to have changed along the twenty years. Although the logistic regression model does not explain the causality between those variables, we can give a valuable interpretation about it: the ones who where active on the labor market in 1993 had somewhat 1.2 times greater propensity to social engagement than the ones who were inactive, whereas, in 2012, the active ones in the workforce show a 25.0% lower

propensity to social engagement than the inactive ones. In other words, twenty years ago, social engagement was a prerogative for workers, nowadays, it seems instead to be related to work-inactivity[4].

Table 1 – *Odd ratio to be active on the labor market, 2012 and 2013.*

Variables	Mod. 1-1993 Exp (ß)	Mod. 2- 2012 Exp (ß)	Mod. 3 – 2012 Exp (ß)
CHRONIC DISEASE (ref. None)			
One	0.682***	0.693***	0.688***
Two or more	0.574***	0.602***	0.601***
RECEIVE MEDICAL CARE (ref. No)			
Yes	1.018	1.145	1.159
SOCIAL PARTICIPATION (ref. Never o Occasionally)			
Frequently	1.215*	0.779**	0.735***
SATISFACTION FOR HEALTH CONDITION (ref. Positive)			
Negative	0.612**	0.736**	0.731**
SATISFACTION FOR ECONOMIC CONDITION (ref. Positive)			
Negative	1.088***	1.252***	1.276**
RESIDENCE LOCATION (ref. South and Islands)			
North-West	0,493***	0,753**	0,724***
North-East	0.549***	0.874	0.847
Center	0.787*	0.875	0.853
EDUCATION (ref. None or Elementary School)			
Middle School	1,212*	1,264*	1,180
High School and over	2,512***	2,504***	2,060***
FAMILY (ref. No children)			
By himself	1,324	1,139	1,160
With his children	1.730***	1.274**	1.273**
AGE (ref. 55-59)			
60-64	0,289***	0,165***	0,169***
65-69	0,063***	0,045***	0,047***
70-75	0.028***	0.019***	0.020***
INTERNET USE (ref. No)			
Yes			1.287**
MOBILE PHONES USE (ref. No)			
Yes			1,421***
Pseudo R^2	42,9%	49,2%	49,6%

*Source: own elaborations based on the data set "Indagine Multiscopo sulle Famiglie", 1993 and 2012***.**

[4] As already mentioned in the introduction of this analysis, that change can be explained by the particular national *policies* that promoted social activeness among older pensioners (Cfr: Davis Smith J., Gay, P., 2005. *Active ageing in active communities*, Bristol, The Policy Press) and the historical events experienced by the cohorts. In fact, the post-war generation reveals different attitudes with respect to the generation who experienced war, as they have lived - albeit from different perspectives - the deepest, most complex and widespread Unionization of the Italian Republican history which, by the way, has created a civic culture transformed only later on in Associations (Mirabile *et al*, 2009).

Besides, two additional issues have been analyzed: formal education - i.e. the second most important variable in the model after "age" - and use of technology as explanatory element of the lifelong learning for older people. In this regard, our analysis detects a greater propensity to activeness among those who have higher educational qualifications. The main explanation of this outcome resides, of course, in the institutional mechanisms: due to longer educational-paths/training-programs, graduated males use to get into the labor market much later and, consequently, develop much later his rights to retire. Furthermore, as regards our hypothesis about the "surfing workers", the ones who have acquired Internet skills tend to remain 1.3 times more active on the labor market, even at old ages (Model 3).

With reference to the socio-territorial determinants, it should be noted that older males who develop a negative perception of their economic status have a greater propensity to activeness than those who consider it quite satisfactory. These outcomes can be read even in terms of active-propensity-differentiations between older males living in the North-West, the most economically developed Italian area, and the ones residing in the South, less developed than the former one. In fact, the propensity of the North-Western older males to remain active is far less than the one shown by the Southern males; yet those differentiation between the two Italian geographical areas seems to have decreased a lot along the two periods of time observed. Then, as family-composition is taken into account, outcomes show that older males keep up active when living with their children: this relationship was more marked in 1993 than it is in 2012. Yet, such data suggests a general lifecycle postponement: as a matter of fact, in Italy, as parents postpone childbearing and their children's postpone housing and economic self-sufficiency, even retirement gets postponed, i.e., parents tend to delay their exit from the workforce.

4. Concluding remarks

The outcomes of this research, about active aging performed by Italian males in 1993 and 2012, reveal that the propensity to remain active on the labour market:

a) slows down as age raises, diminishing more markedly from 60-64 years old onwards;
b) is strongly linked to both health and economic conditions (either real or perceived ones) of older males: as their health gets worse, their propensity to

remain active gets low; still, as their economic situation gets more unreliable, their propensity to remain on the labour market gets higher;

c) is related to different kinds of social engagement, depending on the pre- and post-war generations;
d) is affected by place of residence (older males show to be more active in the Southern than in the Northern) and family setting (older males are generally more active when they live with their children);
e) increases as higher is the education/qualification level held by older males. Nevertheless, propensity to activeness is also marked when technological knowledge is possessed, somewhat, even when no higher education has been accomplished.

This analysis cannot be disregarded from those aspects that inevitably marked differentiations between the two periods observed. Thus, in transition between 1993 and 2012, the propensity to remain active has been strongly affected by the retirement age-raising imposed by the various Italian legislations, the progress had among healthcare and prevention, the national absolute/relative poverty escalation (recently, poverty has much affected the older population, particularly the one living in the Southern), the evolving of "social participation" for the older population, the technological knowledge which is significantly influencing activeness among older people regardless their education level, though it is still nowadays, as it surely was in the past, a decisive active-component.

References

ALA-MUTKA, K. MALANOWSKI N., PUNIE Y., CABRERA, M. 2008. *Active Ageing and Potential of ICT for learning*. Bruxelles: European Commission.

ATTWOOD C., SINGH G., PRIME D., CREASEY R. 2003. *2001 Home Office Citizenship Survey: people, families and communities*. London: Home Office.

CICCARONE G. 2012. *EEO Review: Employment policies to promote active ageing, 2012. Italy*. European Employment Observatory.

COMMISSIONE EUROPEA. 2008. *State of Art of the Active Ageing Policies in Italy: Background of the Italian Legislative Framework*. Fondazione Giacomo Brodolini, vp/2008/006/14.

DAVIS SMITH J., GAY, P. 2005. *Active ageing in active communities*, Bristol, The Policy Press.

INNOCENTI E., VECCHIATO T. (eds) 2013. *Volontariato e invecchiamento attivo*. Pisa: Cesvot edizioni, Quaderno n. 65.
ISTAT, http://dati-censimentopopolazione.istat.it/.
MIRABILE L.M., DE SARIO B., MATSROPIETRO E. 2009. *L'anziano come risorsa*. Roma: IRES Istituto ricerche economiche e sociali.
PEACOCK E. S. 2009. ICT for Employability. In CABRERA M. and MALANOWSKI M. (Eds) *Information and Communication Technologies for Active Ageing*. Bruxelles: EC.
POPULATION REFERENCE BUREAU. 2011. Volunteering and Health for Aging Populations, *Today's Research on Aging*, Vol. 21, pp. 7.
ZUCCHINELLI E., JONES A.M., RICE N., HARRIS A. 2010. The Effects of Health Shocks on Labour Market Exits: Evidence from the HILDA Survey. *Australian Journal of Labour Economics*, Vol. 13(2), pp. 191-218.

SUMMARY

L**ooking Back To Look Forward: The Italian Active Ageing In Between The Old And The New Millennium**

This research is aimed at analyzing the determinants which have influenced the choice of the older (i.e., 55-75 year-old) Italian men to remain active on the labor market at the beginning of active-ageing policies implementations (in 1993) as well as twenty years later on (in 2012). This research takes into account the national representative dataset "*Indagine Multiscopo sulle Famiglie*" (carried on by ISTAT) in order to assess active-ageing determinants through binomial logistic regression. Even though "age" represents the main determinant of activeness, some other important outcomes have emerged as regards health, social participation and use of technologies.

DE PALMA Francesca, Università degli Studi di Bari "A. Moro", franci2910@libero.it.
GIRONE Stefania, Università degli Studi di Bari "A. Moro", stefaniagirone@libero.it.
GRUBANOV-BOŠKOVIĆ Sara, Hebrew University of Jerusalem, saragrubanov@gmail.com

CLUSTER WEIGHTED BETA REGRESSION

Luciano Nieddu, Cecilia Vitiello

1. Introduction

The analysis of data assuming values in the real open interval (0;1) is a common issue in quantitative research when the effect of selected variables on the conditional expectation of a percentage or rate is considered.

In the literature, various alternative methods to model ratios and percentage data have been proposed (see e.g. Papke and Wooldridge, 1996 and Kieschnick and McCullough, 2003). A possible solution is to transform the dependent variable y, for instance using a logit or a probit transformation, so that it assumes values on the whole real line, and then model the mean of the transformed response as a linear predictor based on a set of covariates applying OLS (Demsez Lehn, 1985) to obtain the parameter estimates. This approach, however, has drawbacks, one of them being the fact that the model parameters cannot be easily interpreted in terms of the average of the original outcome but in terms of the transformed response. Furthermore the assumptions of OLS regression are often not met despite the transformation of the data.

An alternative is to use a regression model that assumes that the response variable follows a beta distribution on the interval (0;1), namely $Y|\mathbf{x} \sim \mathcal{B}(p,q)$:

$$f(y;p,q) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} y^{p-1}(1-y)^{q-1}, \quad y \in (0,1),\ p,q > 0, \tag{1}$$

with $\mathbb{E}(Y) = \frac{p}{p+q}$ and $Var(Y) = \frac{pq}{(p+q)^2(p+q+1)}$.

An alternative parameterization, more convenient for modeling purposes, is the one introduced by Ferrari and Cribari-Neto (2004) were the parameters are expressed in terms of the mean μ and the precision parameter φ:

$$\mathbb{E}(Y) = \mu \qquad Var(Y) = \frac{\mu(1-\mu)}{1+\varphi}$$
$$f(y;\mu,\varphi) = \frac{\Gamma(\varphi)}{\Gamma(\mu\varphi)\Gamma((1-\mu)\varphi)} y^{\mu\varphi-1}(1-y)^{(1-\mu)\varphi-1}, \quad y \in (0,1) \tag{2}$$

The variance of a beta-distributed random variable is a scaled version of the binomial variance and the precision parameter allows for a wide range of shapes for the density.

Estimates of the model parameters can conveniently be obtained using maximum likelihood (ML) estimation (Ferrari and Cribari-Neto, 2004)

Classical beta regression models conveniently account for over dispersion by including a precision parameter φ to adjust the conditional variance of the outcome. On the other hand, it is often observed that over dispersion depends on the values of one or more predictor variables (Simas et al., 2010). In this case it is possible to extend the beta regression model by regressing the precision parameter on a subset of the predictor variables.

The aim of this paper is to extend the classical framework of beta regression and suggest a methodology that can help capture unobserved heterogeneity between observations that follow a beta distribution using a cluster weighted modeling approach introduced by Gershenfeld (1997).

In the next Section the proposed model will be introduced and ML estimates for the parameters will be obtained. In Section 3 the model will be applied on a real dataset and in Section 4 some conclusions will be drawn.

2. The Model

Finite mixtures of linear regressions are sometime inadequate for some applications (Hennig, 2000), since they assume *assignment independence*, i.e. the prior probability for single unit (y, $\mathbf{x}$) generated by one of the components of the mixture is constant over all possible values of the vector of covariates $\mathbf{x}$.

Let (y,$\mathbf{x}$) be a set of random variables (a random response variable y and a random vector $\mathbf{x}$) with join density $f(y, \mathbf{x})$. Let's further assume that the support of $\mathbf{x}$ can be partitioned into K subsets.

A more flexible family of mixture models can be obtained assuming that the prior probability for a unit to belong to a cluster depends on the value of the vector of covariates $\mathbf{x}$. This approach was introduced by Gershenfeld (1997) and is known as cluster-weighted models (CWMs), i.e.:

$$f(y, \mathbf{x}) = \sum_{k=1}^{K} \pi_k f(\mathbf{x}|k) g(y|\mathbf{x}, k) \tag{3}$$

where $g(y|\mathbf{x}, k)$ is the conditional density of the response variable given the set of covariates and the group the unit belongs to; $f(\mathbf{x}|k)$ is the distribution of the covariates given the group and π_k is the prior probability of a unit to belong to group k.

Cluster weighted regression models constitutes a flexible family of models to fit the joint density of a set of covariates and a response variable assuming that they are coming from a heterogeneous population.

We will assume that $X|k \sim MVN(\mu_k, \Sigma_k)$ and $Y|X, k \sim \mathcal{B}(\mu'_k, \varphi_k)$.
The location and the dispersion parameter can be linked to the linear predictors as follows:

$g_1(\mu'_k) = \eta_i = \mathbf{x}_i\beta$
$g_2(\varphi_i) = \xi_i = \mathbf{z}_i\gamma$

The functions $g_1(\cdot)$ and $g_2(\cdot)$ are monotonic link functions. Suitable candidates are respectively logit and probit.
The likelihood function for the proposed model is

$$L(\cdot) = \prod_{i=1}^{n} \sum_{k=1}^{K} \pi_k \{ f(y_i|\mathbf{x}_i\beta_k\gamma_k) g(\mathbf{x}_i|\mu_k, \Sigma_k) \} \tag{4}$$

let $\theta_k = (\beta_k, \gamma_k)$ and $\omega_k = (\mu_k, \Sigma_k)$, *k=1,..., K* .

ML equations for the parameters of the Beta model:

$$\begin{cases} \dfrac{\delta \ell(\cdot)}{\delta \theta_k} = \sum_{i=1}^{n} w_{i,k} \dfrac{\delta log(f(y_i|\mathbf{x}_i\theta_k))}{\delta \theta_k} = 0 \\ w_{ik} = \dfrac{\pi_k f(y_i|\theta_k) g(\mathbf{x}_i|\omega_k)}{\sum_{k=1}^{K} \pi_k f(y_i|\theta_k) g(\mathbf{x}_i|\omega_k)} \end{cases}$$

and ML equations for the parameters of the Gaussian process:

$$\begin{cases} \dfrac{\delta \ell(\cdot)}{\delta \omega_k} = \sum_{i=1}^{n} w_{i,k} \dfrac{\delta log(g(x_i|\omega_k))}{\delta \omega_k} = 0 \\ \\ w_{ik} = \dfrac{\pi_k f(y_i|\theta_k) g(\mathbf{x}_i|\omega_k)}{\sum_{k=1}^{K} \pi_k f(y_i|\theta_k) g(\mathbf{x}_i|\omega_k)} \end{cases}$$

are both weighted score equations with weights given by the a posterior probabilities $w_{i,k}$ of unit *i* to belong to component *k.*
This yields to standard results for the estimates of the parameters μ_k and Σ_k:

$$\hat{\mu}_k = \frac{\sum_{i=1}^{n} \mathbf{x}_i w_{ik}}{\sum_{i=1}^{n} w_{ik}} \quad \hat{\Sigma}_k = \frac{\sum_{i=1}^{n} (\mathbf{x}_i - \bar{\mathbf{x}}_k)(\mathbf{x}_i - \bar{\mathbf{x}}_k)^T w_{ik}}{\sum_{i=1}^{n} w_{ik}}$$

while estimates for the a priori probabilities can be obtained solving the following constrained ML problem:

$$\begin{cases} \dfrac{\delta \ell(\cdot)}{\delta \pi_k} = \sum_{i=1}^{n} \dfrac{f(y_i|\theta_k) g(\mathbf{x}_i|\omega_k)}{\sum_{k=1}^{K} \pi_k f(y_i|\theta_k) g(\mathbf{x}_i|\omega_k)} + \lambda = 0 \\ \\ \sum_{k=1}^{K} \pi_k = 1 \end{cases}$$

yielding: $\widehat{\pi_k} = \sum_{i=1}^{n} w_{i,k}$.

3. Real data example

The U.S. News data contains information on tuition, room and board costs, SAT or ACT scores, application/acceptance rates, graduation rate, student/faculty ratio, spending per student, and a number of other variables for a total of 35 categorical and quantitative variables over a sample of more than 1300 schools.

The dataset is taken from the 1995 U.S. News & World Report's Guide to America's Best Colleges and is freely available from the statlib repository (http://lib.stat.cmu.edu/datasets/colleges/). Most of the data are for the 1993-94 school year. Two third of the schools are private (65.19%).

The rate of accepted applicants has been considered as response variable and "instate tuition" (X_1) and "sfratio" (student/faculty ratio" X_2) have been used as covariates. Only records with no missing data have been considered.

The proposed model has been fit to the data using BIC to select the optimal number of components. The best value of BIC was obtained in correspondence of K=5.

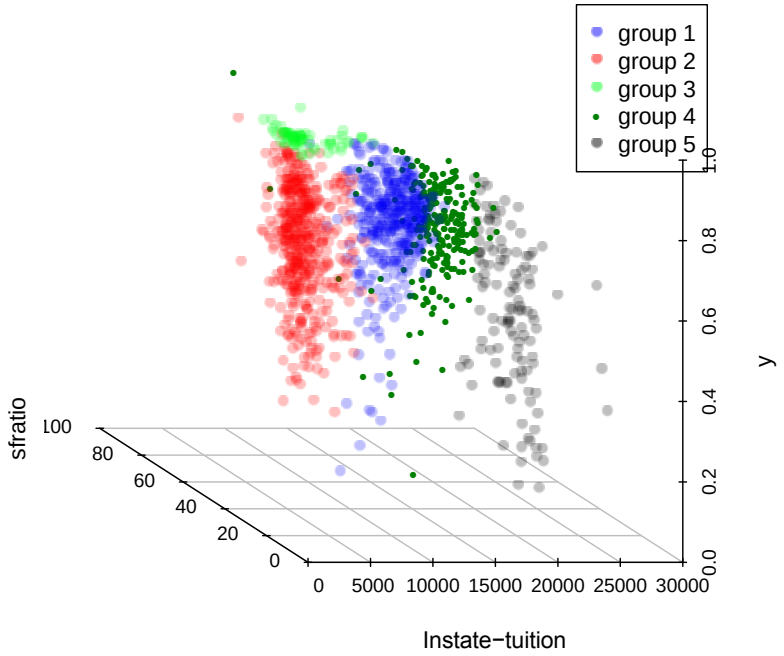
The results of the estimates for location and precision models for the Beta distribution have been reported in Table 1. Only the values that, at a confidence level of $\alpha = 0.05$, were significantly different from zero have been retained.

Table 1 – *MLE for cluster weighted Beta regression model with K=5 components.*

Group	Estimates	Intercept	X1	X2
1	β	0.4944	0.0001	n.s.
	γ	-1.8959	0.0005	n.s.
2	β	0.8280	n.s.	n.s.
	γ	3.5407	-0.0001	-0.0557
3	β	n.s.	n.s.	0.1212
	γ	n.s.	n.s.	n.s.
4	β	0.9844	n.s.	n.s.
	γ	n.s.	0.0003	n.s.
5	β	n.s.	-0.0001	0.1031
	γ	n.s.	n.s.	n.s.

In Figure 1 a 3d-plot of the 5 groups has been displayed to easy the interpretation of the results while in Table 2 the distribution of the schools by group and type (public/private) has been reported.

Figure 1 – *Cluster structure with respect to acceptance rate, instate-tuition and sfratio.*



Considering the results for the location parameter, instate-tuition is influential for the rate of acceptance in groups 1 and 5, while student-to-faculty ratio is only influential in Group 3. Group 1 and 5 are mainly private schools (Table 2). Group 5 is made of very expensive and very well known universities, and shows negative coefficient for the variable "instate-tuition": for those famous highly qualified universities, high tuitions means being able to apply a very strict selection of the applicants. In Group 1 we find mainly private colleges with different vocations, for them an increase in tuition increases the proportion of accepted applicants. Therefore instate-tuition are effective for these two groups, representing for one a measure of the selectiveness of the university (Group 5) for the other (Group 1) a measure of the quality of the college (private colleges with very small instate-tuitions could be considered just a way to get a degree).

Variable X_2 is influential only on groups 3 and 5. Student-to-faculty ratio can be considered a structural variable indicating the dimension of the school and therefore its capacity to accept students. An increase of X_2 in Group 5 increases the

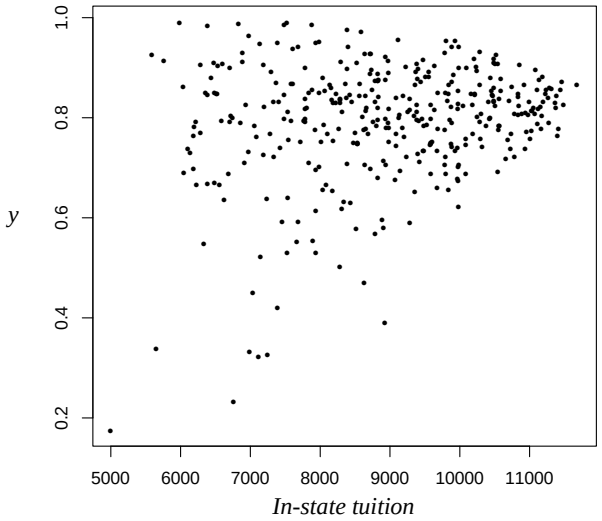
proportion of accepted applicants (the university accepts students up to its structural capacity identified by the predefined student to faculty ratio set by the board of directors). This is valid also for Group 3, where it is the only variable affecting the proportion of applicants. Group 3 is a highly heterogeneous group of schools (Table 2) with very low instate-tuitions and a very high acceptance rate (Figure 1).

Table 2 – *Distribution of schools by group and type (public/ private).*

Group	Public	Private	**Total**
1	1	370	**371**
2	389	63	**452**
3	48	20	**68**
4	1	244	**245**
5	0	125	**125**
Total	**439**	**822**	**1261**

Let's now consider the effect of those variables on the precision parameter. To better understand the effect of the precision parameter let's consider Group 1 where the precision parameter depends only on instate-tuition and its effect on the precision of the estimate of the response variable may be shown in a two-dimensional plot. In Figure 2 a plot of the outcome variable and the instate-tuition for all colleges of group 1 has been displayed.

Figure 2 – *Effect of covariates on precision: instate-tuition vs acceptance rate*



The triangular shape of the cloud implies that there is large variability of the outcome y for small values of the covariate and the variability decreases as the covariate "instate tuition" increases. Therefore the precision of the estimates for *y* is greater for higher values of instate-tuition. Group 2 and Group 4 have two diverging behaviors: they both have an admission rate which does not depends on X_1 and X_2, with the admission rate of Group 4 being slightly higher than that of Group 2. In Group 4 the precision of the estimates increases with instate-tuitions and does not depend on $X_{2,}$ while in Group 2 it decreases with instate-tuition and with X_2.

4. Conclusions

We have proposed a Beta regression model based on CWRs that allows for flexibility on modeling both the location and the precision parameter for the beta distribution. Our proposal, which should include the finite-mixture approach as a particular case, not only can be used in presence of over dispersed data but it can also be used as a diagnostic tool to detect a mixture structure in the data. The proposed methodology has been tested on benchmark data yielding very interesting results.

Riferimenti bibliografici

DEMSETZ H., LEHN K. 1985. The structure of corporate ownership: causes and consequences. *Journal of Political Economy*, 93, 1155–77.

FERRARI S.L.P., CRIBARI-NETO F. 2004. Beta regression for modeling rates and proportions. *Journal of Applied Statistics* 31: 799–815.

GERSHENFELD N. 1997. Nonlinear inference and Cluster-Weighted Modeling. *Annals of the New York Academy of Sciences*, 808(1), 18–24.

HENNIG C. 2000. Identifiablity of models for clusterwise linear regression. *Journal of Classification* 17, 273–296.

KIESCHNICK R., MCCULLOUGH B.D. 2003. Regression analysis of variates observed on (0, 1): Percentages, proportions and fractions. *Statistical Modelling* 3: 193– 213.

PAPKE L.E., WOOLDRIDGE J.M. 1996. Econometric methods for fractional response variables with an application to 401(k) plan participation rates. *Journal of Applied Econometrics* 11: 619–632.

SIMAS A.B., BARRETO-SOUZA W., ROCHA A.V. 2010. Improved estimators for a general class of beta regression models. *Computational Statistics & Data Analysis* 54: 348–366.

SUMMARY

Cluster Weighted Beta Regression

Beta regression is the standard method to explore how a response assuming values in (0;1) depends on a set of covariates. With respect to standard regression, in this case, the parametric model requires two systems of equations: one for the mean and the other for the precision parameter that can be based on the same set of covariates.

Therefore for two different sets of covariates and the same value of the linear predictor for the mean we could have different precisions.

Nevertheless a linear model for the precision parameter could not be good enough to capture all the heterogeneity in the data.

We will extend the characteristic approach of cluster weighted linear models to the beta regression problem in order to obtain a flexible model both in analyzing relations between means and covariates and in evaluating prediction precision.

Luciano NIEDDU, UNINT, l.nieddu@unint.eu
Cecilia VITIELLO, Department of Statistics, "Sapienza" University of Rome, cecilia.vitiello@uniroma1.it

Rivista Italiana di Economia Demografia e Statistica Volume LXVIII n. 3/4 Luglio-Dicembre 2014

POVERTÀ ED ESCLUSIONE SOCIALE DELLE FAMIGLIE IN ITALIA[1]

Domenica Quartuccio, Giorgia Capacci

1. Il concetto di benessere e la fonte dei dati

Il concetto di benessere è stato storicamente dibattuto, in letteratura, tra chi pone l'attenzione sulle dimensioni (uni o multidimensionale), sulle tipologie di variabili (soggettive, oggettive o entrambe), sull'aspetto individuale o su quello collettivo, e così via. Le teorie sul benessere possono essere, in generale, racchiuse in due grandi gruppi. Nel primo rientrano quelle elaborate negli anni Sessanta che focalizzano l'attenzione sull'individuo e sulla critica al benessere inteso solo in senso materiale ed economico. Tali teorie, sviluppate nell'ambito di un filone di studi denominato "qualità della vita", vedono tra i maggiori sostenitori studiosi come W. Zapf, il quale dichiarò che il benessere è formato da una componente soggettiva e da una oggettiva e può essere descritto in maniera trasversale rispetto ai diversi domini sociali (Zapf W., 1999). Dall'incrocio delle dimensioni soggettivo/oggettivo e buono/cattivo derivano le seguenti tipologie: benessere (buono su tutti e due gli assi) e deprivazione (cattivo su entrambi gli assi), adattamento (buono soggettivo), dissonanza (cattivo soggettivo).

Un altro famoso studioso di queste teorie è stato Amartya Sen, secondo il quale il benessere è inteso come sviluppo di *capabilities* necessarie per raggiungere i funzionamenti a cui gli individui attribuiscono alto valore (Sen A., 2000). La misurazione del benessere, quindi, non è più tarata sui livelli effettivamente raggiunti, ma sul fatto che ogni individuo disponga di determinate capacità personali che gli permettano di raggiungere gli obiettivi di benessere e di qualità della vita a cui aspira.

Nel secondo filone di studi sul benessere, di cui Susanna Terracina è una degli esponenti, invece, si pone l'accento sugli aspetti relazionali e distributivi; si ritrovano in questo ambito gli studi sulla coesione sociale, l'inclusione e l'esclusione sociale e lo sviluppo umano.

[1] Seppure il saggio vada considerato come frutto di una comune riflessione di entrambe le autrici, i paragrafi 1 e 2.1 sono stati redatti da Giorgia Capacci, i paragrafi 2.2 e 3 da Domenica Quartuccio.

Nasce in questo periodo anche una questione rilevante nella misurazione del benessere, cioè se sia necessario far riferimento alla dimensione individuale o a quella familiare. "Il benessere individuale si può costruire anche (o soprattutto) in ambito familiare […] per cui sarebbe errato considerare gli individui senza considerare il contesto in cui vivono. Per questo gli studi sulla povertà considerano sempre più spesso la dimensione familiare [...]. Il dibattito sull'utilizzo di indicatori di tipo individuale e familiare è comunque sempre aperto e in continua evoluzione […]" (Terracina S., 2003).

Obiettivo di questo lavoro è fornire una panoramica sul benessere, e quindi sul malessere, delle famiglie residenti in Italia, ovvero analizzare il legame che esiste tra le strutture familiari (e caratteristiche socio-demografiche dei suoi componenti) e le condizioni di povertà e di esclusione sociale. La fonte utilizzata è l'indagine Istat "Reddito e condizioni di vita", denominata Eu-Silc (European Union Statistics on Income and Living Conditions), che fornisce, a livello europeo, le statistiche ufficiali su povertà, benessere e condizioni economiche delle famiglie. Il nostro studio ha focalizzato l'attenzione sull'indicatore sintetico di *rischio di povertà o di esclusione sociale* che rientra nella strategia *Europa 2020*. Tale indicatore, fornito proprio da Eu-Silc, è usato per monitorare uno degli obiettivi della strategia, quello della "lotta alla povertà e all'emarginazione" attraverso l'uscita di almeno 20 milioni di persone da questa condizione entro il 2020.

Si definiscono a rischio di povertà o di esclusione sociale le famiglie che sperimentano almeno una delle seguenti tre condizioni: 'a rischio di povertà', cioè con un reddito equivalente inferiore al 60% del reddito mediano; 'a bassa intensità di lavoro', famiglie i cui componenti di età 18-59 anni lavorano meno di un quinto del tempo che avrebbero potuto lavorare nel periodo di riferimento; in 'condizioni di severa deprivazione materiale', ovvero in una situazione di involontaria incapacità di sostenere spese per determinati beni o servizi[2].

I dati utilizzati nel presente studio sono quelli dell'indagine condotta dall'Istat nel 2012 su un campione di 19.579 famiglie (47.365 individui).

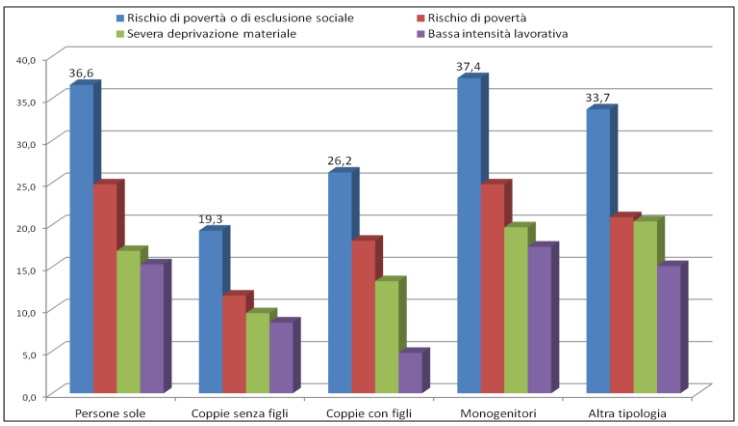
[2] Si tratta di famiglie che registrano almeno quattro (su una lista di nove) segnali di deprivazione materiale: non poter sostenere una spesa imprevista, non potersi permettere di fare una settimana di ferie all'anno, un pasto adeguato ogni due giorni, di riscaldare adeguatamente l'abitazione, essere in arretrato con il pagamento del mutuo o dell'affitto, delle bollette o di altri debiti, non potersi permettere lavatrice, tv, telefono o automobile.

2. Analisi dei dati

2.1. Analisi descrittiva delle famiglie

Nel 2012, le famiglie residenti in Italia che sperimentano una condizione di povertà o di esclusione sociale sono quasi un terzo del totale (29,2%) (Istat, 2013). Il livello di vulnerabilità cambia in funzione delle differenti tipologie familiari: l'Istat misura che quelle più esposte al rischio sono i monogenitori (37,4%) e le persone sole (36,6%) (Figura 1). Considerando le coppie con figli (26,2%), se ve ne sono tre o più, l'indicatore sale ancora arrivando al 38,6%. Di contro, le coppie senza figli sembrano stare meglio, infatti l'indicatore si attesta al 19,3%.

Figura 1 – *Famiglie secondo gli indicatori di deprivazione per tipologia familiare, Anno 2012.*



Fonte: Elaborazioni su dati Eu-Silc.

Per individuare le categorie di famiglie maggiormente esposte a condizioni di malessere, abbiamo analizzato i dati facendo riferimento alle caratteristiche socio-demografiche del principale percettore di reddito della famiglia.

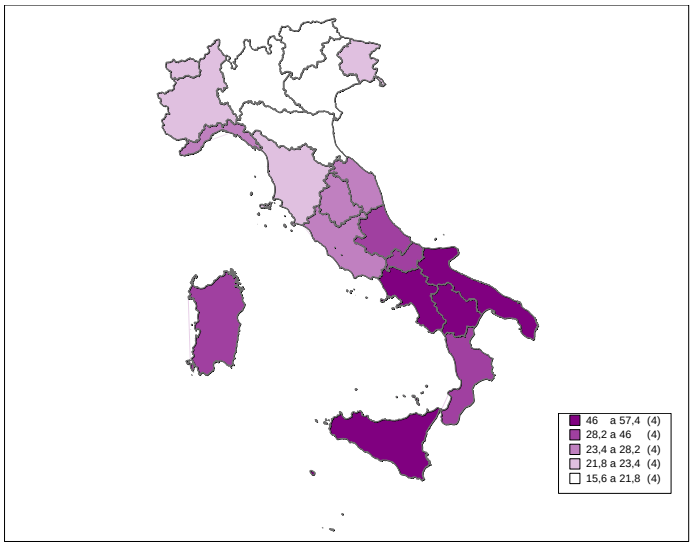
Dal punto di vista del genere, le famiglie in cui il percettore è una donna sono generalmente più esposte al rischio di povertà o di esclusione sociale di quelle in cui è un uomo, a conferma delle condizioni di maggiore difficoltà in cui esse versano.

Relativamente al livello di istruzione, i dati confermano quanto ci si aspetta, ovvero il possesso di un titolo di studio superiore tende a preservare dal rischio di povertà o di esclusione sociale. Le famiglie più vulnerabili, infatti, sono quelle in cui il principale percettore di reddito ha al massimo la licenza elementare (41,4%),

mentre, al contrario, se egli ha conseguito un titolo studio pari o superiore alla laurea il rischio si riduce fortemente (circa il 13%).

La presenza nel mercato del lavoro del principale percettore di reddito della famiglia influisce fortemente sulle condizioni di malessere/benessere. Le famiglie in cui è disoccupato hanno un valore dell'indice sintetico molto alto, pari al 72,8%; se, invece, egli è ritirato dal lavoro[3] la percentuale si attesta al 55,5%. Naturalmente le condizioni migliorano, e di molto, se le famiglie hanno come percettore principale un occupato: in particolare, va meglio se l'entrata familiare principale proviene da un lavoratore dipendente (20,7%) piuttosto che da un autonomo (26,0%).

Figura 2 – *Famiglie a rischio di povertà o di esclusione sociale per regione, Anno 2012.*



Fonte: Elaborazioni su dati Eu-Silc.

Infine, l'enorme disuguaglianza territoriale è confermata dall'andamento dell'indicatore sintetico di rischio di povertà o di esclusione sociale: la percentuale di famiglie vulnerabili che risiede al Mezzogiorno è più che doppia rispetto a quella di chi risiede al Nord (rispettivamente 46,8% e 19,3%). Proseguendo l'analisi a livello regionale, è la Sicilia a detenere il record negativo. In questa regione, infatti, le famiglie a rischio di povertà o di esclusione sociale sono quasi il doppio rispetto alla media nazionale (57,4% contro il 29,2%) (Figura 2). Situazione

[3] Il ritirato dal lavoro è chi ha cessato un'attività lavorativa per raggiunti limiti di età, invalidità o altra causa; la figura del ritirato dal lavoro non coincide necessariamente con quella del pensionato in quanto, non sempre, il ritirato dal lavoro gode di una pensione.

allarmante si trova anche in Puglia (48,6%) e in Campania (47,9%); mentre le regioni più virtuose sono l'Emilia-Romagna (15,6%), il Veneto (16,3%) e la Lombardia (19,9%). L'eccezione a questo dualismo territoriale è sicuramente la Liguria, regione del Nord che presenta un indice di rischio piuttosto alto (23,4%).

2.2. Analisi multivariata: modello di regressione logistica

Per analizzare i dati anche con un approccio multivariato abbiamo scelto di applicare un modello di regressione logistica scegliendo come variabile dipendente l'indicatore di rischio di povertà o di esclusione sociale degli individui ('a rischio di povertà o di esclusione sociale' vs. 'non a rischio di povertà o di esclusione sociale') e come variabili indipendenti le seguenti:

- ripartizione geografica di residenza;
- quinti di reddito;
- tipologia familiare;
- titolo di godimento dell'abitazione;
- variabili demo-sociali relative al principale percettore di reddito della famiglia (sesso, classe di età, titolo di studio e condizione professionale).

Dall'analisi dei risultati si evince che, in riferimento alla tipologia familiare, le persone sole con meno di 65 anni e gli individui in famiglie monogenitori sono le categorie più esposte al rischio di povertà o di esclusione sociale; al contrario, sono meno esposti gli individui che vivono in famiglie di coppie con figli (Tavola 1). Le persone anziane sembrano, in qualche modo, essere tutelate dal rischio di povertà e di esclusione sociale: le coppie senza figli in cui la persona di riferimento ha più di 65 anni e i single della stessa fascia di età, infatti, sono molto meno vulnerabili rispetto alla categoria di riferimento. Questo fa pensare che, probabilmente, la persona anziana che ha un'entrata fissa, come potrebbe essere la pensione, riesce a tutelare la propria famiglia e a non essere più un peso, quanto piuttosto una fonte preziosa di aiuto. Tale teoria è confermata anche guardando l'età del principale percettore di reddito della famiglia: chi lamenta peggiori condizioni sono gli individui che vivono in famiglie in cui il principale percettore ha meno di 35 anni di età. In Italia, quindi, donne e giovani sono il tassello più debole del sistema.

Per quanto riguarda la condizione professionale, si conferma quanto visto con l'analisi descrittiva: chi vive in famiglie in cui il principale percettore è un lavoratore dipendente o autonomo è molto meno a rischio rispetto a chi, invece, è disoccupato o, peggio ancora, si trova nella condizione di "altro non occupato"[4].

Prendendo in considerazione un indicatore strettamente monetario come il reddito familiare, misurato in quinti, si vede come la vulnerabilità delle persone

[4] Appartengono a questa categoria gli studenti, le casalinghe, gli inabili al lavoro, eccetera.

residenti in Italia sia fortemente legata alla disponibilità economica delle famiglie in cui vivono: al crescere del reddito decresce, infatti, la condizione di rischio.

Tabella 1 - *Rischio di povertà o di esclusione sociale per caratteristiche familiari e del principale percettore, modello di regressione logistica: parametri, standard error e significatività - Anno 2012.*

Caratteristiche del principale percettore e familiari		Coefficienti beta	Standard error	Significatività (a)
Quinti di reddito (ref=1°)	2°	-2,70	0,04	***
	3°	-3,84	0,05	***
	4°	-4,36	0,06	***
	5°	-4,56	0,06	***
Sesso (ref=Uomini)	Donne	0,09	0,04	
Classi di età (ref= Fino a 34 anni)	35-44 anni	-0,43	0,05	***
	45-54 anni	-0,35	0,05	***
	55-64 anni	-0,04	0,06	
	65 anni e più	-0,75	0,08	***
Livello di istruzione (ref= Fino a secondaria inf.)	Secondaria superiore	-0,54	0,03	***
	Universitaria e oltre	-0,71	0,06	***
Condizione professionale (ref= Dipendente)	Autonomo	0,17	0,05	**
	Disoccupato	1,58	0,07	***
	Ritirato dal lavoro	1,25	0,06	***
	Altro non occupato	1,65	0,06	***
Tipologia familiare (ref= Coppie senza figli con P.R. (b) meno di 65 anni)	Coppie senza figli con P.R. di 65 anni e più	-0,75	0,08	***
	Coppie con figli	-0,15	0,06	**
	Monogenitori	0,66	0,07	***
	Persone sole meno di 65 anni	0,86	0,07	***
	Persone sole 65 anni e più	0,27	0,08	**
	Altra tipologia	0,16	0,08	*
Ripartizione geografica (ref= Nord)	Centro	0,40	0,04	***
	Sud e Isole	0,64	0,03	***
Titolo di godimento dell'abitazione (ref=Proprietà)	Affitto	-0,30	0,04	***
	Usufrutto e uso gratuito	0,23	0,05	***

(a) *ρ≤0.05 **ρ≤0.01 ***ρ≤0.001
(b) P.R.: Persona di Riferimento

Chi appartiene al quinto più ricco, quindi, è molto meno esposto rispetto a chi si trova nel quinto più povero.

Anche l'analisi multivariata, inoltre, conferma l'enorme disagio del Mezzogiorno nei confronti del Nord, in primis, ma anche del Centro, in termini di povertà e deprivazione.

In un'ottica di genere, viene confermato che gli individui che vivono in famiglie in cui il principale percettore di reddito è una donna sono più a rischio rispetto a quelle in cui è un uomo. Guardando al livello di istruzione si evidenzia che il rischio di povertà o di esclusione sociale degli individui decresce al crescere del titolo di studio conseguito dal principale percettore di reddito della famiglia.

Dato che la maggioranza delle famiglie vive in case di proprietà, ci è sembrato opportuno osservare anche gli effetti sul modello di una variabile come il titolo di godimento dell'abitazione: così come ci si aspetta, gli individui che vivono in famiglie 'in affitto' o 'in usufrutto/uso gratuito' sono più vulnerabili rispetto a quelle proprietarie.

3. Conclusioni e prospettive future

La profonda crisi economica che ha, e ancora sta attraversando, l'Italia in questi ultimi anni è registrata anche dai dati dell'indagine Eu-Silc. Considerando gli anni che vanno dal 2010 al 2012, ovvero quelli in cui la crisi ha avuto gli effetti più devastanti nella nostra economia, l'indicatore fin qui analizzato ha segnato un incremento notevole, passando dal 24,5% del 2010 al 28,2% del 2011, per poi salire ancora al 29,9% nel 2012.

Essendo l'indagine Eu-Silc condotta nei paesi dell'Unione Europea, obiettivo futuro del lavoro sarà quello di effettuare un confronto internazionale dei dati al fine di analizzare eventuali aggregazioni territoriali fra paesi in termini di povertà e deprivazione per costruire una mappa dell'indicatore a livello europeo.

Ulteriori sviluppi saranno quelli di verificare se i risultati ottenuti dal modello di analisi multivariata di regressione logistica sono confermati impiegando anche altri strumenti di analisi esplorativa, come, ad esempio, analisi in componenti principali e cluster analysis oppure analizzare l'indicatore in una serie storica degli ultimi dieci anni (l'indagine Eu-Silc si svolge a livello europeo dal 2004) per valutare quali variabili hanno inciso maggiormente ad una sua crescita, tenuto conto anche degli effetti della crisi economica in atto.

Ringraziamenti

Si ringrazia la dott.ssa Lucia Coppola, ricercatore dell'Istat, per i preziosi suggerimenti forniti nella preparazione del lavoro e nell'analisi dei risultati.

Riferimenti bibliografici

ANNONI P., WEZIAK-BIALOWOLSKA D. 2012. Quality of Life at the sub-national level: an operational example for the EU, *JRC Scientific and Policy Reports*.

BOLASCO S. 1999. *Analisi muldimensionale dei dati*. Roma: Carocci.

ISTAT. 2014. *Rapporto Annuale 2014. La situazione del paese*. Roma: Istat.

ISTAT. 2013. *Reddito e condizioni di vita.* Comunicato stampa. Roma: Istat, http://www.istat.it.

SEN A. 2000. *Lo Sviluppo è libertà*. Milano: Mondadori.

TERRACINA S. 2003. Dimensioni e indicatori sociali dello sviluppo. *Contributi di ricerca IRES-Piemonte*. 172/2003

ZAPF W. 1999. Social Reporting in the 1970s and 1990s. *Working paper*, FS III 99-404.

SUMMARY

Poverty and Social Exclusion in Italy

Aim of this paper is to analyze the risk of poverty or social exclusion in relation to different household types. The results of the applied logistic regression model show that households where woman is the main income earner are those with higher risks as well as couples with children and young couples. Households living in Southern Italy have a higher risk of poverty than those living in the North, the only exception is Liguria. Data used for this research are those of the last available Eu-Silc survey (2012) carried out for European countries.

Domenica QUARTUCCIO, ISTAT, quartucc@istat.it
Giorgia CAPACCI, ISTAT, gcapacci@istat.it

Rivista Italiana di Economia Demografia e Statistica Volume LXVIII n. 3/4 Luglio-Dicembre 2014

A MODEL BASED CATEGORISATION OF THE ITALIAN MUNICIPALITIES BASED ON NON-RESPONSE PROPENSITY IN THE 2011 CENSUS

Antonella Bernardini, Andrea Fasulo, Marco D. Terribili

1. Introduction

The counting operations carried out during a population census can be afflicted by non-sampling errors.

The quality takes on the meaning of precision that is expressed as an inverse function of the statistical error. The aim of the Istat is to provide accurate estimates of the main non-sampling errors, particularly in complex investigations like the Census. The non-sampling error is a function of many factors: organizational aspects of the survey, the behaviour of a plurality of individuals or Institutions.

The Italian National Institute of Statistics (Istat) certifies the quality of the 15th Population and housing census through a sample survey of coverage assessment, as required by Commission Regulation (EU) No 1151/2010 of 8th December 2010 implementing Regulation (EC) No 763/2008 of the European Parliament and of the Council. The Post Enumeration Survey (PES) has the goal of estimating the real number of the people living in Italy on 9 October 2011, at the reference day of the 15^{th} population and housing general census; it has also the aim of evaluating the errors of overcoverage and undercoverage in the individuals count.

The main indicators to evaluate the accuracy, is the coverage rate, which is calculated (under the assumption of to not undercover the population) as the ratio between the number of the enumerated units during the census and the real population dimension, denoted by N and obviously unknown.

The survey design of the PES is a two stages with stratification of the primary sample units (252 municipalities) and of the secondary units (about 2500 enumeration areas). The collection of data has been planned to guarantee the independence between the two surveys. The interest of the survey is focused on the families and on the individuals habitually living in the enumeration areas selected for the sample of the PES.

In order to estimate the coverage rate we have estimated a statistic model based on the Petersen's model assumption; this model is part of a models class, called dual-system (or capture-recapture methods) and it represents one of the most common model between those used to quantify the Census coverage errors (Wolter, 1986). One of the basic hypothesis of the estimation model used is the constant capture probabilities at the census and at the PES, for all the units belonging to the soubpopulation.

We need to fit the estimation model to small domains in which the capture probability is the same and then to calculate the estimate in wider domains, given by aggregation of sub-domains. In estimation phase, thanks to a greater number of auxiliary available variables, regarding design and sampling phase, a post-stratification has been carried out.

One of the used post-stratification variables is the Hard To Count index (HTC), which contributes to detect homogeneous areas relatively to the difficulty of a subpopulation to be enumerated. The model study, on which the index has been designed, leads to analyse social, economic and demographic characteristics, significantly influential on the individual probability to be censused. These characteristics point out some differences, relatively to local non-response levels.

Following the important ONS experience about the HTC applied during the population census of 2001 and 2011, an index has been studied to categorize Italian municipalities regarding an homogeneous expected level of right enumeration of the individuals.

2. Predictive models for right enumeration

To study the propensity of the individuals to be correctly numbered during the Population Census, data coherence with the Post Enumeration Survey (PES) has been taken into account. With the aim of output the individual estimated probability of right enumeration, a predictive model has been fitted; this model assumes a link function between several auxiliary variables, collected during the PES or available from other sources, and the dependent variable. The latter is a binary variable that points out the missing record linkage between the individuals listed during the Post Enumeration Survey and those ones listed during the Population Census. So the variable modalities are:

$$Y = \begin{cases} 1 & \text{unsuccessfull record linkage} \\ 0 & \text{successful record linkage} \end{cases}$$

Being the dependent variable a binary one, the implemented models are fixed effects logistic ones, they can be expressed in the following way:

$$\text{Logit } P(Y_i = 1 | X_i) = \text{Log } \frac{P(Y_i = 1 | X_i)}{1 - P(Y_i = 1 | X_i)} = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + ... + \beta_k X_{ki}$$

As an alternative to fixed effects models, Random-Effect Logit Models are implemented too, to take into account the enumeration areas (territorial division in which Municipalities are divided) with the intercept γ_d :

$$\text{Logit } P(Y_{id} = 1 | X_{id}) = \text{Log } \frac{P(Y_{id} = 1 | X_{id})}{1 - P(Y_{id} = 1 | X_{id})} = \alpha + \beta_1 X_{1id} + \beta_2 X_{2id} + ... + \beta_k X_{kid} + \gamma_d$$

Auxiliary variables, available for the statistical units reached by the Post Enumeration Survey, describe socio-demographic characteristics of the individual and of the municipalities/provinces of which they belong to. Post-stratification allows to exploit the data richness of the Post Enumeration Survey, its updated individual information, and to integrate it with other local variable, available from archive.

Table 1 − *Auxiliary variables, regarding informative level*

Level	Auxiliary variable
Individual	Age
	Age classes
	Sex
	One unit family
	Extended family (more than 7 individuals)
	Foreigners
	Singles (Separated, divorce, widow)
	Proxy student (19≤age≤30, educational qualification at least diploma
Municipal	University city
	Coastal city
	Altimetric zones (in 5 modalities)
	Population density (pop. Per km^2)
	Foreigners rate
Provincial	Unemployment rate
Interactions	Foreigners * Foreigners rate
	One unit family * Age class 10÷29
	University city * Proxy student

In the model study phase three alternative models have been proposed: the first one fits only individuals variables, the second model fits area variables in addition to the individual ones and the third fits also interactions between some variables paired. In the following table 1, the complete list of auxiliary variables, distinct for degree of detail and other.

3. Hard To Count model

The multi-level modeling involves the prediction of the variance at different levels, so often it start with an analysis to determine what levels this variation can be considered significant. In the first step two random intercepts were tested, one at the municipal level and one at enumeration area level, because it is useful to assess how much of the total variance is explained between the different groups. This can be accomplished by calculating the Intraclass Correlation Coefficient (ICC) using the formula:

$$ICC = \tau_{00}/(\tau_{00} + \sigma^2) \quad (1)$$

where is the between-group or Intercept variance, and the within-group or residual variance. The estimated ICC, at the municipal level, is .009, while at the enumeration area level is .032, a value that makes us lean towards that level of detail. In the second and last step, the significance of the Intercept variance was evaluated through a likelihood ratio test. In order to do this we compare the values of -2 log likelihood of the null model with random intercept with the likelihood of the null model without random intercept. The value of - 2 log likelihood for the model without the random intercept is -579.870. The same indicator for the model with the random intercept is -584.294. The difference of 4.423 is significant for a chi-square distribution with one degree of freedom. These results suggest that a random intercept of enumeration area produces a significant improvement of the model. It has been estimated that 3.2% of the total variance in the study of non-response probability, is a function of the enumeration area of the person.

Even the study of the model was performed in different phases. The model selected was made through the use of commonly used criteria for the choice of models that are the log-likelihood, the AIC and BIC indicators. In the first phase, the variables of the questionnaire, available for each person, have been used. The best was the model with the variables age classes, sex and citizenship, with AIC, BIC and log-likelihood respectively equal to 29.381, and 29.466 -14.682. Afterwards, area level covariates were added and the best model was the one with the variable rate of unemployment, university common flag, population density and rate of foreigners. Adding area level covariates, led an improvement of all 3 indicators, which amounted to 29.196 AIC, 29.324 BIC and -14.586 log-likelihood.

Finally, were considered the combined effects of different variables, but the only significant interaction, which improved the model was between citizenship and the rate of foreign residents in the municipality. Also adding this effect the AIC, BIC, and the log-likelihood are equal to 29.174, 29.313 and -14.574.

Table 2 shows the regression coefficients for the three models described above.

Table 2 − *Regression coefficients of the models.*
In grey the coefficients not significant

Auxiliary variables	Individual variables model	Individual + area level variables model	Complete model
Intercept	-5,711	-6,905	-7,067
Age class 10-29	0,075	0,074	0,072
Age class 30-49	0,048	0,046	0,041
Age class 50-74	-0,555	-0,555	-0,564
Age class ≥75	-0,481	-0,480	-0,488
Sex (female)	-0,164	-0,166	-0,168
Foreigners	2,395	2,395	2,848
Unemployment rate		10,411	10,489
University city		0,826	0,826
Population density		9,505e-05	9,178e-05
Foreigners rate		4,594	6,817
Foreigner * Foreigners rate			-5,795

Once calculated the probability of being been counted or not at the census, these were averaged at the municipal level, so as to return to the spatial detail of interest.

The orderly distribution of the predicted values, relative to the 252 municipalities of the sample, was divided on the basis of percentiles in 3 modes following the distribution 40% - 40% - 20%. Thus the virtuous municipalities, with a low problem with counting the person, will be categorized with the HTC level 1, the municipalities in an intermediate situation, will have the HTC level 2, and the most problematic municipalities from the point of view of the correct enumeration will have the HTC level 3. This categorization has also been applied to probability

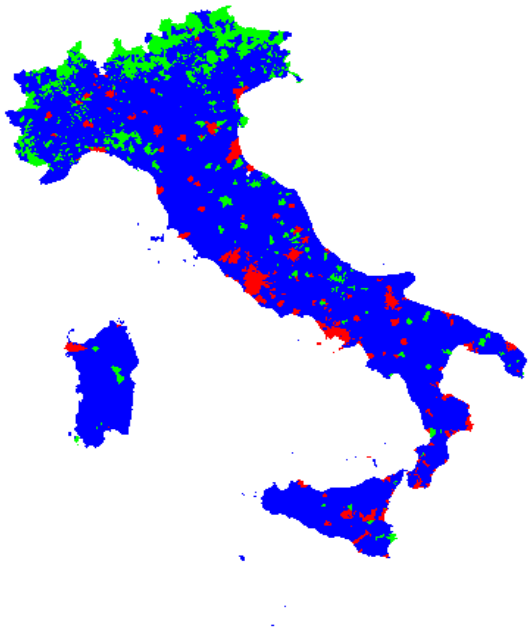
of the municipalities outside the sample, predicted by using only the synthetic part of the multilevel logistic regression model described above.

4. Results

The available wealth of information has allowed a detailed study on the hardest individuals to count in the census.

Figure 1 shows the distribution of HTC among Italian municipalities.

Figure 1 − *HTC distribution in the Italian Municipalities.*
HTC level 1 in green, HTC level 2 in blue, HTC level 3 in red.



The most virtuous municipalities, colored in green, are those that are distributed along the Alps and Apennines, show small municipalities. Municipalities with an intermediate index, colored in blue, are the majority and they cover almost the entire territory. Finally, the most problematic areas are colored red and representing large municipalities, focusing long the Italian coast highlighting the issues related to the second home or holiday house and movements for seasonal work.

References

Abbott O. 2000. 2001 Hard to Count Index, *One number census steering committee*. http://www.ons.gov.uk/ons/guide-method/census/census-2001/design-and-conduct/the-one-number-census/methodology/steering-committee/key-papers/hard-to-count-index.pdf

Grossi P., Mazziotta M. 2012. *Qualità del 15° Censimento generale della popolazione e delle abitazioni attraverso una indagine di controllo che misuri il livello di copertura*. Istat Working Papers n. 16/2012

Office for National Statistics. 2011. Office for National Statistics, London

Predicting patterns of household non response in the 2011 Census. http://www.ons.gov.uk/ons/guide-method/census/2011/the-2011-census/processing-the-information/statistical-methodology/predicting-patterns-of-household-non-response-in-the-2011-census.pdf

SUMMARY

A model based categorisation of the Italian municipalities based on non-response propensity in the 2011 Census

The Italian National Statistical Institute had certified the quality of the 15th Italian population and housing census thanks to a Post Enumeration Survey (PES) taken throughout the months immediately after the Census. The aim of the PES is to produce total estimates adjusted for under coverage and, for the first time, over coverage.

The model underlying the under and over coverage estimation, takes into account the differences between individual probabilities of responding to the Census. For this aim a regression unit-level model was applied; in order to study the individual probability to be censused on the basis of which the Hard to Count Index (HTC) of Italian municipalitiesit was created. In the model were used variables derived from the PES questionnaire and additional area-level variables from other sources.

HTC categorises the 8092 Italian municipalities in 3 different levels, partitioning the distribution of municipal non-response propensities, based on percentiles.

This paper describes in detail the multilevel logistic regression model used to study non-response probability, the development of the HTC, the methods and the analysis carried out to evaluate the goodness of index, regarding the census coverage.

Antonella BERNARDINI, Italian National Institute of Statistics, anbernar@istat.it
Andrea FASULO, Italian National Institute of Statistics, fasulo@istat.it
Marco D. TERRIBILI, Italian National Institute of Statistics, terribili@istat.it

Rivista Italiana di Economia Demografia e Statistica Volume LXVIII n. 3/4 Luglio-Dicembre 2014

SPATIAL ANALYSIS OF EMPLOYMENT MULTILPLIERS IN SPANISH LABOR MARKETS

Margherita Gerolimetto, Stefano Magrini

1. Introduction

Governments all over the world, in particular of those countries where the Global Financial Crisis has hitted most intensively, allocate increasing amounts of public financial resources to promote economic development and foster employment. These economic development policies are very often implemented at a local level, in response to bottom up and decentralizing strategies.

In general, the main target of local development initiatives, which mainly consist of promoting new businesses' start-up or favouring the development of existing businesses, is the so-called tradable sector, i.e., the sector producing goods that tend to be sold and consumed elsewhere. This attention to the tradable sector is essentially motivated by the alleged presence of a "multiplier" effect that benefits the entire (local) economy: an initial "injection" might generate a larger final impact on the local economy due to the emergence of a positive, cumulative effect arising from the links between the tradable sector and the rest of the economy.

There are several different approaches to estimate the magnitude of the multipliers. Quite often, the quantification of multiplier effects is carried out through Input-Output tables making a distinction between three types of effects: direct, indirect and induced. The direct effect is the effect that takes place in the targeted industry; the indirect effects concern inter-industry transactions: as a new firm opens, it will demand locally produced materials, possibly resulting in a further creation of jobs; induced effects measure the effects of the changes in overall household income brought in by the employment increase. Within this context, the employment multiplier is then represented by the ratio of direct plus indirect plus induced employment changes to the direct employment change.

In a rather recent paper, however, Moretti (2010) argues that this way of quantifying the multiplier effect might be inaccurate as it tends to overlook offsetting general equilibrium effects on local prices: as employment increases also wages (unless local labour supply is infinitely elastic) and land rents (unless land supply is infinitely elastic) do, thus imposing cost increases to all local firms. As a consequence, the author suggests an alternative method for estimating the size of the multiplier effect based on a simple regression framework. Apart from allowing for general equilibrium effects, this method has the additional important advantage of being particularly easy to implement, thus providing analysts and policy makers with an easy-to-use tool to evaluate the consequences of policy actions.

The aim of this work is to analyse, using the empirical methodology proposed by Moretti as a basis, the presence and magnitude of local multipliers within Spanish local labour market areas. The paper is structures as follows: in the second Section the local nontradable multiplier is presented, the third Section describes the data set, the fourth Section is devoted to the empirical analysis and the conclusive comments.

2. Local non tradable multiplier

Moretti's (2010) aim is to estimate the long term employment multiplier at the local level and presents a simple, informal, spatial equilibrium framework, according to which a positive shock to a tradable industry has: *i*) a positive effect on employment both in the nontradable sector and in other tradable industries and *ii*) offsetting general equilibrium effects (due to the increase in wages and land rents).

To do this, Moretti extends the simple spatial equilibrium model by Rosen and Roback (Rosen, 1979, Roback, 1982) featuring: production of both tradable and nontradable goods; non-homogeneous labour in tastes and skills; imperfect labour mobility due to idiosynchratic location preferences.

In Moretti's (2010) conceptual framework, policy intervention might attract new firms or increases product demand for existing firms. The local nontradable multiplier in a city is

$$M = \frac{\Delta N^T + \Delta N^{NT}}{\Delta N^T} = 1 + \frac{\Delta N^{NT}}{\Delta N^T}$$

where ΔN^T is the tradable sector labor change and ΔN^{NT} is the nontradable sector

labor change. Moretti estimates the elasticity $\varepsilon = \frac{\Delta N^{NT}}{\Delta N^{T}} \times \frac{N^{T}}{N^{NT}}$ and then focusses on $\frac{N^{T}}{N^{NT}}$ calculated as $\hat{\varepsilon} \times \frac{N^{NT}}{N^{T}}$. Operatively, the elastiticy is estimated via a simple linear regression

$$g^{NT} = \mathrm{a} + \mathrm{b}g^{T} + u \quad (1)$$

where g is the growth rate (change in the log number) of jobs and u is the usual iid error term.

As anticipated, the main advantage in using this approach is that it overcomes the tendency of the traditional methodology, i.e. local Input-Output, to overlook the offsetting general equilibrium effects (Moretti, 2010). In addition, as emphasised by de Blasio and Menon (2011), with the present methodology the exogenous variation is directly attributed to the tradable sector which in fact attracts most of the politcy interventions.

There are however two potentially critical issues. The first is the existence of spatial dependence, implied, for example, by trade and migration flows that lead to feedback across cities. This issue is totally neglected by Moretti despite a large literature suggesting that untreated spatial dependence might affect the estimates (among others, LeSage and Pace, 2009). To deal with this, the literature offers a list of modelling strategies, among which the most common are the spatial lag and the spatial error models.

The second critical issue is represented by a possible inverse causation of the variables in the regression that, in turn, implies endogeneity. Moretti (2010) treats this by adopting the instrumental variables estimator (IV) where the istrument is represented by the potential growth rate that each labor market area would have experienced had its economic subsectors grown at the corresponding national average growth rate. This is a rather commonly used instrument (Bartik, 1991). In his empirical analysis, Moretti (2010) obtaines that an additional job in the tradable sector leads to 2.77 (OLS) or 1.59 (IV) in the nontradable one.

Here, we deal with both issues simulateously proposing a tentative apprach in two steps. Firstly, we remove spatial dependence through a spatial filter; then, we proceed by estimating via IV. More formally, we estimate via IV

$$g_F^{NT} = \mathrm{a}' + \mathrm{b}' g_F^{T} + u' \quad (2)$$

where

$$g_F^{NT} = \left(I - \rho_{NT} W\right)^{-1} g^{NT}$$

$$g_F^{T} = \left(I - \rho_{T} W\right)^{-1} g^{T}$$

where W is a row-standardized spatial weight matrix.

3. Data

The empirical analysis considers a sample of 103 Spanish Local Labour Market areas (LLMAs). The data, collected by the Spanish Ministry of Empoyment and Social Security, gather the quarterly occupational statistics in the 60 CNAE (Clasificación Nacional de Actividades Económicas, the Spanish adaptation to the NACE classification) economic subsectors, which have been recorded for each local labour market area from 1999 to 2012. The Social Security database includes, for each municipality, observations for each of the 60 economic subsectors of the CNAE classification. To build the dataset, data at the municipality level have been gathered together according to the definition of the LLMAs.

In extreme synthesis, the reasons why LLMAs are preferred to administrative regions are two. Firstly, the choice of the territorial unit must ensure that the effects of local policies are confined, as far as possible, within the targeted area (Cheshire and Hay, 1989; Cheshire and Magrini, 2006). Secondly, results depend on size and shape of spatial units in an apparently unpredictable way thus leading to what is called Modifiable Areal Unit Problem (MAUP). In this framework, Openshaw (1996) claims that the MAUP will disappear once researchers know what the areal objects they wish to study are and according to Arbia (1989, 1991) it becomes essential to use units characterised by 'significant boundaries' from an economic standpoint.

The variables of interest to estimate model (1) are obtained by grouping employment data for 60 subsectors. The traditional approach to separate tradable from nontradable jobs, adopted in the studies by Moretti (2010) and de Blasio and Menon (2011) identifies tradable industries with manufacturing and nontradable ones with services. However, as emphasised by Jensen and Kletzer (2005) many service activities can in fact be considered as producing tradable goods and some manufacturing goods can be included among the nontradables. Consequently, rather that adopting the traditional classication, we use the two-digit code classification provided by Hufbauer and Vieiro (2013) (based on the approach developed by Jensen and Kletzer, 2005). According to this approach, when production is concentrated at a distance from consumption within the US, as inferable from a locational Gini coefficient exceeding 0.1, the activity is classified as tradable.

4. Empirical analysis

The aim of the current regression analysis is to estimate the elasticity of nontradable employment with respect to tradable. In particular, we first conduct a traditional OLS analysis; then, we proceed with an IV estimate to address endogeneity concerns; finally, we repeat the IV estimate on previously spatially filtered data in order to deal, simultaneously, with both spatial dependence and potential endogenity issues. All results are reported in Table 1.

Table 1 – *Estimates.*

	OLS		**IV**		**Spatial Filter + IV**	
	Coeff (s.d.)	p-value	Coeff (s.d.)	p-value	Coeff (s.d.)	p-value
Intercept	0.3151 (0.0316)	0.000	0.5033 (0.082)	0.000	0.480 (0.067)	0.000
Tradable	0.5036 (0.080)	0.000	-0.0104 (0.226)	0.963	0.039 (0.180)	0.829
	Statistic	p-value	Statistic	p-value	Statistic	p-value
Moran's *I*	4.766	0.000	3.9554	0.000	-1.169	0.242

As shown in the second and third column of Table 1, the OLS estimates of the b coefficient in model (1) are significant. However, they are not reliable, as we know that endogenity is likely to affect them. Consequently, to tackle this issue, the IV estimates are computed and this clearly changes the results, since now the coefficient of the tradable is not significant anymore (fourth and fifth column). The last row of Table 1 displays the results of the Moran's I test of spatial dependence; spatial patterns are significantly found both in the OLS and IV residuals and this leads to the last step of our analyis which is represented by an IV regression of the spatially filtered varibales, whose outcomes are reported in the last two columns of Table 1. The spatial filter is carried out using a maximum likelihood estimator of ρ_{NT} and ρ_T, W is a 15 neighbors row-stardardized spatial weight matrix. The results confirm that also once both issues are taken into account, the coefficient multiplying the growth rate of the employment in the tradable sector is not significant.

Along the lines of Moretti (2010) we computed (when possible) the job effect, i.e. the jobs in the nontradable sector created by an additional job in the tradable one and present in Table 2 we present a comparison of the results obtained in this work with those obtained by Moretti (2010) and De Blasio and Menon (2011).

Table 2 – *Estimated elasticities and additional non tradable jobs for each additional tradable job.*

	Spain		**Italy**		**US**	
	Coeff	Job Effect	Coeff	Job Effect	Coeff	Job Effect
OLS	0.504	0.671	0.061	n.a.	0.554	2.77
IV	0	0	0	0	0.335	1.59
Sp. Filter + IV	0	0				

All in all, once spatial dependence and endogeneity are simultaneously accounted for, the estimated coefficient is no longer statistically significant thus leading to the conclusion that the analysis of the Spanish Local Labour Market Areas does not provide evidence in support of local multipliers. In other words, once spatial spillover effects and endogeneity are controlled for, the relationship between the growth rate of the employment in the Tradable Sector and the one recorded in the NonTradable sector does not reveal any multiplicative effect. What emerges from the Spanish case, as well as from the Italian study by De Blasio and Menon, is therefore a rather different picture with respect to the US case. A possible motivation for this might be that Spanish and Italian labour markets are not as flexible as the US labour markets; in addition, it is likely that labour and land supplies are less elastic than in the US (due to lower job and geographical mobility) thus leading to stronger offsetting general equilibrium effects.

Further work could possibly proceed along the following direction. First, the model should include other variables, next to the growth rate of the tradable sector – that are more context-specific, and affect the magnitude of local multiplicative effects. These factors could be grouped into two main categories depending on how they affect the final effect. A first group directly affects the local labour market, by modifying the elasticity of the local labour demand or supply. For example, the high degree of centralization of the wage determination, the lower propensity to

move from the hometown to main job-attracting cities, and a more structured welfare system make wages less responsive to the local economic conditions and the local economy will experience a lower degree of job turnover.

References

ARBIA G., 1989, *Spatial Data Configuration in Statistical Analysis of Regional Economic and Related Problems*. Kluwer, London.

ARBIA G., 2001, Modelling the Geography of Economic Activities in a Continuous Space, *Papers in Regional Sciences*, 80, 411-424.

BARTIK T.J., 1991, *Who Benefits from State and Local Economic Development Policies?* W.E. Upjohn Institute for Employment Research. Kalamazoo, MI.

de BLASIO G., MENON C., 2011, Local Effects of Manufacturing Employment Growth in Italy. *Giornale degli Economisti e Annali di Economia*, 70, 101-112.

CHESHIRE P.C., HAY, D.G, 1989, *Urban Problems in Western Europe: an economic analysis*. Unwin Hyman, London.

CHESHIRE P.C., MAGRINI S., 2006, Population Growth in European Cities: Weather Matters – But only Nationally. *Regional Studies*, 40, 23-37.

HUFBAUER G.C., VIEIRO M., 2013, *Corporate Taxation and US MNCs: Ensuring a Competitive Economy*. Policy Brief 13-9, Washington, DC: Peterson Institute for International Economics.

JENSEN J.B., KLETZER L., 2005, *Tradable services: understanding the scope and impact of services outsourcing*. Working paper series, n. 05-9. Institute of International Economics. Washington, DC.

LESAGE J.P., PACE, R.K., 2009, *Introduction to Spatial Econometrics*. Chapman and Hall, Boca Raton, FL..

MORETTI E, 2010, Local Multipliers, *American Economic Review: Papers and Proceedings*, 100, 1-7.

OPENSHAW S., 1996, *Developing GIS-Relevant Zone-Based Spatial Analysis Methods*, In Longley P. e Batty M. (Eds.) Spatial Analysis: Modelling in a GIS Environment, Cambridge: GeoInformation International.

ROBACK J., 1982, Wages, Rents and the Quality of Life. *Journal of Political Economy*, 90, 1257-1278.

ROSEN S., 1979, *Wage-based Indexes of Urban Quality of Life*. In Miezkowski, P. and Straszheim, M.R. (eds.) Current Issues in Urban Economics. Johns Hopkins University Press. Baltimore, MD.

SUMMARY

Spatial analysis of employment multilpliers in spanish labor markets

The purpose of this work is to investigate the effect of employment promotion policies in Spain. In general, this depends on the ability of the intervention at creating new jobs in the targeted area, but also, to a large extent, on the impact they have on other parts of the local economy. Estimating the latter effect of the local multiplier, is therefore, important for regional economic development policies. Along the lines of Moretti (2010), we present an empirical analysis of the local multipliers using data of Spanish labor market areas over the period 1999-2007. From the methodological point of view, in this work not only endogenity (via instrumental variables estimates), but also spatial effects are taken into account. According to the results, the inclusion of spatial effects reveals the magnitude of the multiplier could be limited.

Margherita GEROLIMETTO, Lecturer in Economics Statistics, University of Venice

Stefano MAGRINI, Professor in Political Economics, University of Venice

Rivista Italiana di Economia Demografia e Statistica Volume LXVIII n. 3/4 Luglio-Dicembre 2014

IMMIGRANTS LIVING IN THE EU15 COUNTRIES AND THEIR CONDITIONS OF INTEGRATION IN THE LABOUR MARKET

Anna Di Bartolomeo, Salvatore Strozza

1. Introduction

The integration of labour migrants and their descendants in the labour market is today a main challenge of European host societies. It has significant and positive implications not only on individuals and their personal autonomy but also on the society as a whole, in terms of social cohesion (OECD, 2009). More integrated migrants are likely to be more productive and, in turn, better accepted by receiving societies. Given the constant need of migrants in today's European labour markets (Fargues, 2011), investing resources on the integration of migrants and social cohesion would thus represent a valid tool in allowing policymakers for designing rationale admission policies, accepted by European civil societies.

This paper aims at depicting the conditions of migrants in the labour market of EU15 countries (EU15) and comparing their situation with that of non-migrants. Different categories of migrants (units of analysis) were built according to the intersection of the following variables: country of birth, country of citizenship, sex and country of destination. Special emphasis is put on the conditions of migrants residing in the EU15 Mediterranean countries, namely France, Greece, Italy and Spain.

2. Labour market indicators and population categories

Integration is an equal and fair participation of migrants in the host societies and, as such, can be empirically defined as a process of convergence between migrants and non-migrants' outcomes, according to a number of dimensions, namely labour market, education, civic and political participation, etc.

Consequently, when approaching integration studies from a quantitative viewpoint, two preliminary issues need to be addressed. First, one should clearly indicate both the dimension of interest and the list of indicators used to measure it. Second, being integration a "convergence" process, two populations should be clearly defined: a) the population who aims at being integrated (migrant population) and b) the population towards which such convergence would occur.

In this paragraph, these issues are discussed while the dimension of integration analysed, its indicators and the population categories here adopted are presented.

In order to identify the dimension to be analysed and its indicators, the Declaration of Zaragoza[1] was used as a benchmark. Among several dimensions (EU, 2010, p.13), this work focuses on the integration of migrants in the labour market given the importance of economic migration in EU Mediterranean countries: today – and despite the on-going global economic recession – these states are still the main receivers of labour migration flows within the EU[2]. Regarding the choice of indicators, we added to the Zaragoza core indicators – employment, unemployment and activity rates (EU, 2010, p.15) – two additional ones: the over-qualification and the self-employment rate[3]. Finally, the distribution of employed migrants by sector (NACE) and by level of occupation (ISCO) were also taken into account.

To define the migrant population, for the sake of comparability, neither the country of citizenship nor the country of birth criterion alone were employed. Indeed, using the foreign population to proxy migrants would have been challenged by very different laws on acquisitions of citizenship between EU Member States[4]. Similarly, considering the foreign born population would have included a portion of people who are not a direct consequence of foreign immigration *stricto senso*. This applies to those countries which have been recently evolved from emigration to immigration countries (Greece or Italy) or in such States with a long colonial history, such as France. Here, the foreign born population also includes emigrants or expatriates' descendants who return to their parents' country of origin or – in the case of France – the so-called repatriates from former colonies (*repatriés*), i.e. categories who are not a direct consequence of immigration but of emigration (Strozza, 2010).

Facing these issues, we decided to classify the population of interest according to the intersection of these two criteria: 1) population holding the citizenship of the country of residence and born in the country of residence (NATNAT), group which corresponds to the non-migrant population and that is used as reference category; 2) population holding the citizenship of the country of residence and born abroad (NATABR), group which include various categories, among which naturalized persons, repatriates from former colonies, children of emigrants; 3) foreign population

[1] The "Zaragoza Declaration", developed within the 4th European Ministerial Conference on Integration (Zaragoza, 15-16 April, 2010) is – at the time of writing – the reference document concerning the key aspects on immigrants' integration in the EU.

[2] Of 3.2 million first residence permits for work reasons granted by EU Member States in the period 2008-2012, more than a half (1.6 millions) were granted by Mediterranean countries (Italy, 1.1 million; Spain, 438 thousand; France, 95 thousand; and Greece, 48 thousand).

[3] The over-qualification rate is defined as the "share of persons with tertiary education working in a low- or medium-skilled job among employed persons having achieved tertiary education" (Eurostat, 2011), while the self-employment rate as the "share of self-employed persons among all employed".

[4] For an overview on EU Member States' citizenship laws, see the "EUDO Observatory of CITIZENSHIP" at http://eudo-citizenship.eu/.

born in the country of residence (FORNAT), category which includes second generation migrants (but only those who have not acquired the citizenship of the host country at birth or subsequently); 4) foreign population born abroad (FORABR), i.e. first generation migrants. Having very different personal backgrounds and social, economic and cultural resources, these population categories are expected to follow very heterogeneous integration trajectories. Moreover, due to the very different integration approaches and migration histories (together with selective admission mechanisms towards first generation migrants) of EU15 countries, the same category is expected to perform differently between countries. The construction of this classification is thus justified by the expected high degree of variability – in terms of integration outcomes – both between and within countries.

3. Data and methods

Data were taken from the 2011 European Union Labour Force Survey (EU-LFS), through which the indicators of integration in the labour market were built for the 4 above-mentioned categories. Such indicators were constructed separately for men and women aged 25-54. This age group was chose as it allows for minimizing the effect of migration related to non-economic reasons, as e.g. study and retirement; as well as the effect of the very different age structures of the national/native-born and the foreign/foreign-born populations. As a matter of fact, it represents a more homogeneous population group, useful for comparison purposes (Eurostat, 2011).

In order to describe the conditions of migrant and non-migrant groups in the labour market in the 14 countries of interest[5], the following raw indicators distributed by sex and population categories were built (for a total of 112 statistical units): employment rate, unemployment rate, activity rate, over-qualification rate, self-employment rate, the distribution of the employed population by sector (NACE) and by occupational level (ISCO).

Concerning the methodology, we adopted a multivariate approach including a principal component and a cluster analysis, the latter implemented on the factorial loadings of statistical units as resulted from the extracted components. These explorative analyses, which are a significant step towards measuring integration levels, were implemented with the aim of identifying similarities and differences between population categories and EU15 States.

[5] Finland is not included in the analysis, because of data unavailability.

4. The profile of migrants in the labour market

The principal component analysis, performed on 112 statistical units *per* 14 indicators, allows us for retaining 4 components with eigenvalues greater than one, that explain almost 75% of the total variance (table 1).

The first component portrays performances – in terms of employment and unemployment rate – and used competencies in the labour market: categories with high employment rates and employed in highly skilled occupations are opposed to categories characterized by high rates of unemployment and over-qualification who are mainly employed in low skilled jobs, household services, accommodation and food service activities. The second component synthetizes labour market participation and employment sectors: categories with high shares of people employed in agriculture, manufacturing and construction activities are opposed to categories who are mainly employed in the tertiary sector, e.g. education, health, public administration, etc. The third and fourth component – of more difficult interpretation – are positively correlated with high percentages of people employed in medium skilled jobs and agriculture, respectively.

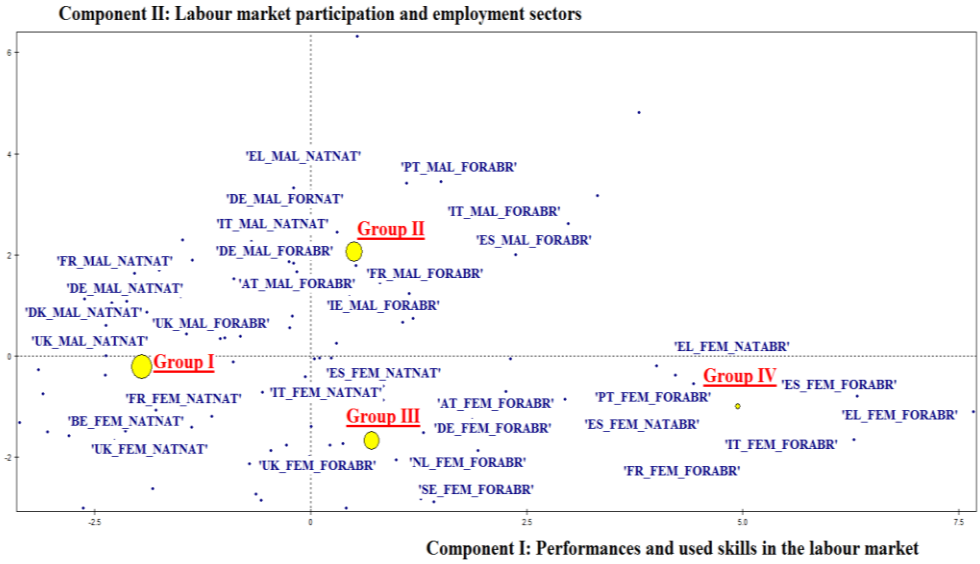
Table 1 – *Correlation (factorial loadings) between raw indicators of labour market and the first 4 factors of the Principal Component Analysis.*

Raw indicators	Factor 1	Factor 2	Factor 3	Factor 4
Employment rate	**-0.74**	0.38	-0.39	-0.21
Unemployment rate	**0.66**	0.14	-0.08	0.33
Activity rate	-0.51	**0.58**	-0.52	-0.08
Over-qualification rate	**0.76**	0.08	-0.24	0.13
Self-employment rate	-0.13	**0.62**	0.27	0.47
% employed in agriculture (NACE A)	0.29	0.43	-0.05	**0.67**
% employed in mining and quarrying, manufacturing and others (NACE B, C, D, E, G)	-0.07	**0.78**	0.28	-0.17
% employed in construction (NACE F)	-0.10	**0.78**	-0.27	-0.07
% employed in accommodation and food service activities (NACE I)	**0.63**	-0.04	-0.20	-0.38
% employed in household services (NACE T)	**0.68**	-0.19	-0.24	0.04
% employed in public administration, education and other tertiary activities (NACE H, J-S)	-0.43	**-0.85**	0.14	0.15
% employed in high-skilled jobs[(a)]	**-0.85**	-0.26	-0.19	0.27
% employed in medium-skilled jobs [(b)]	0.36	0.45	**0.67**	-0.29
% employed in low-skilled jobs [(c)]	**0.78**	-0.10	-0.42	-0.07
% explained variance	31.77	23.74	10.66	8.61

Note: (a) High-skilled jobs include: legislators, senior officials and managers, professionals, technicians and associate professionals. (b) Medium-skilled jobs include: plant and machine operators and assemblers, craft and related workers, skill agricultural and fishery workers, service workers and shop and market sales workers, clerks. (c) Low-skilled jobs include elementary occupations.

The cluster analysis was then performed through hierarchical aggregating methods by using, as variables, the loadings of the 112 statistical units observed in the 4 retained components. Such analysis identified 4 groups, clustering categories with similar conditions in labour market insertion dynamics (figure 1).

Figure 1 – *Factorial plans: Cluster Analysis results*



Group I includes the "most and best inserted" categories (for a total of 40 categories), being characterized by high labour market participation (high rates of activity) and good performances (high employment and low unemployment rates). It is also distinguished by low over-qualification rates, high shares of people employed in highly skilled jobs in the tertiary sector and, meanwhile, low shares of individuals employed in medium-low skilled jobs in the household services, accommodation and food service activities. Concerning categories, group I is mainly composed of males and females of non-migrant groups (NATNAT) residing in all countries except of three Mediterranean states (Greece, Italy and Spain) together with Portugal and Ireland, i.e. all EU15 countries with a recent immigration history. The most significant trait of this group seems the "non-migrant origin", though there are some migrant categories, too. For instance, it includes second generation migrants (FORNAT) born in Denmark, Luxembourg and Sweden as well as French, Dutch, UK male nationals born abroad (NATABR), who are likely to be the result of their colonial past and, finally, Portuguese people born abroad, resulting from return migration patterns of emigrants' children. In addition to these "selected" categories, it is worth noting as this group also includes

some categories of foreigners born abroad (FORABR). This is the case of male foreigners born abroad and residing in the UK, who certainly result from the highly selective admission policies put in place by UK governments in the last 25 years (Di Bartolomeo and Fargues, 2014).

Group II includes categories with "high participation but difficult insertion in the labour market" (33 categories). It is indeed characterized by high activity rates and high levels of unemployment denoting such groups that are characterized by a large availability of labour supply for whom, however, the insertion in the market is difficult and burdensome. These categories are mainly employed in low qualified jobs in the agricultural, manufacturing and construction sectors. Last but not least, the rate of self-employment is high. This cluster is strongly gender characterised: almost all categories are composed of males[6]. Regarding their origin, group II includes all categories of foreigners born abroad (FORABR) with some few exceptions[7] as well as some second generation migrants (FORNAT) residing in Western (Austria, Belgium and Germany) and Southern (Greece, Italy, Spain and Portugal) EU countries. Again, it is worth noting as some non-migrant groups (NATNAT) belong to group II, namely those residing in Mediterranean countries (with the exception of France) and in Ireland shedding light on how labour market conditions are still (and perhaps even more following the on-going global economic recession) difficult here for both migrant and non-migrant populations.

Group III comprehends those categories with "low labour market participation and strong concentration in the tertiary sector" (31 categories), being positively correlated with people employed in education, Public Administration, services, etc. and negatively with the activity rate, the employment rate and the share of people employed in manufacturing and construction sectors. The gender component is fundamental here as well: almost all groups are composed of females. As with the origin, this group includes foreigners born abroad (FORABR) residing in Northern and Western EU countries. In addition, like their male counterparts, female non-migrant populations (NATNAT) of two Mediterranean countries – Italy and Spain – are not found in the "most and best inserted" group (group I) but here, i.e. the group which comprehends the majority of female migrants (group III).

Group IV includes categories with "low participation and difficult insertion in the labour market, mainly concentrated in household private services" (8 categories). It is not only characterized by low activity rates and high unemployment levels, but also by high share of people employed in household services, accommodation and food service activities. However, the fact that these sectors are characterized by a high presence of informal activities suggests that a so low attitude to enter the labour

[6] Exceptions are non-migrant females (NATNAT) in Greece and Portugal and female foreigners born in Italy and Portugal (FORNAT).

[7] With the exception of those residing in Luxembourg, Sweden and UK.

market may hide some unobservable dynamics. Moreover, this cluster is composed of categories mainly employed in low skilled jobs with high levels of over-qualification. It seems thus to identify the most disadvantaged group, which stay at the borders of the labour market or, at least, of the informal one. In this group, again, only females are found and specifically, female foreigners born abroad (FORABR) and residing in Mediterranean countries of more ancient (France) and recent (Greece, Italy and Spain) immigration together with Portugal.

5. Conclusions and further direction of the research

Our analysis finds that labour market insertion's modalities largely differ according to migrant categories, as defined by the intersection of country of birth, country of citizenship, sex and country of residence. In particular, the variable sex is determinant in creating two separate groups within the migrant population. In addition, creating migrant categories according to the combination of the criteria of country of citizenship and country of birth has been a correct strategy. Indeed, while foreigners born abroad (FORABR) and people holding the citizenship and born in the country of residence (NATNAT) are two defined and separated categories, the population holding the citizenship of the country of residence and born abroad (NATABR) as well as the foreign population born in the country of residence (FORNAT) perform very differently. Moreover, it is worth noting as even some non-migrant groups do belong to different clusters suggesting as the non-migrant population residing in Mediterranean countries perform worse than their counterpart living in other EU15 countries. The latter is indeed entirely found in the "best performing" cluster (group I), while the former is found in two other clusters, where migrant groups are overrepresented: males belong to the cluster characterized by "high participation but difficult insertion in the labour market" (group II) while females to the group with "low labour market participation and strong concentration in the tertiary sector" (group III).

Eventually, these differential outcomes suggest that in order to measure integration trajectories, it is necessary to consider the relative rather than the absolute position of migrants with respect to non-migrant groups. So, regarding the future directions of the research, once selected the most adequate indicators (e.g. employment and unemployment rates, over-qualification and self-employment rate) – which, if necessary will be reoriented so that the higher the value of indicators, the higher the level of integration – they will be relativized with respect to the average situation of non-migrant populations of each country (the majority group), so that the different conditions of national labour markets are taken into account. The synthesis through a unique composite index will allow us for ranking different migrant categories by labour market integration and for further evaluating the correlation with migration and integration policies put in place by each country of analysis.

Acknowledgments

This work has been developed within the framework of the research project "Dimensions, measures and determinants of immigrants' integration in host societies" [*Dimensioni, misure e determinanti dell'integrazione degli immigrati nelle società di destinazione*] (directed by Prof. Salvatore Strozza) co-financed by *Compagnia San Paolo*, *Banco di Napoli* and *Polo delle Scienze Umane e Sociali* of University of Naples Federico II, programme FARO 2010/2011 (CUP: E61J12000180005).

References

DI BARTOLOMEO A., FARGUES P., 2014. The Geography of Highly-Skilled Arab Migration, in FARGUES P., VENTURINI A. (Eds) *Arab migrants and the global competition for talents*, Londra: IB Tauris.

EUROSTAT, 2011. *Migrants in Europe 2011 edition. A statistical portrait of the first and second generation Migrants in Europe*, Paris: European Commission.

FARGUES P., 2011. *International Migration and Europe's Demographic Challenge*, EU-US Immigration Systems 2011/09, Robert Schuman Centre for Advanced Studies, San Domenico di Fiesole (FI): European University Institute.

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT (OECD), 2009. *The labour market integration of immigrants and their children. Key findings from OECD country reviews*, High-Level Policy Forum on Migration, Paris, 29-30 June 2009, Paris: OECD.

STROZZA S., 2010. International migration in Europe in the first decade of the 21st century, *Rivista Italiana di Economia Demografia e Statistica*, Vol. LXIV, No. 3, pp. 7-43.

SUMMARY

This note aims at depicting the conditions of migrants in the labour market of EU15 countries and comparing their situation with that of non-migrants. By adopting a multivariate approach, our analysis finds that labour market insertion's modalities largely differ according to migrant categories, as defined by the intersection of country of birth, country of citizenship, sex and country of residence.

Anna DI BARTOLOMEO, Migration Policy Centre of EUI, anna.dibartolomeo@eui.eu
Salvatore STROZZA, University of Naples Federico II, strozza@unina.it

Rivista Italiana di Economia Demografia e Statistica Volume LXVIII n. 3/4 Luglio-Dicembre 2014

MACHINE LEARNING AND TEXT MINING TO CLASSIFY TWEETS ON A POLITICAL LEADER

Agostino Di Ciaccio, Giovanni Maria Giorgi

1. Introduction

The Social Network Twitter was created in 2006, but it has had a slow expansion in Italy starting from 2009. Twitter is now very popular and counts 255 million users, becoming the social media most used by public personalities, showmen, politicians. In Twitter, each user handles his own personal page that can be updated via text messages, with a maximum length of 140 characters, known as "*tweets*". Anyway, the user can add links to pictures, videos, or other documents. The limit on the length of each tweet is, at the same time, the strength and weakness of this social network: with 140 characters you cannot develop a speech, but you can write a sentence quickly using a smartphone.

Let us recall some of the unique aspects of this social network. A Twitter user can choose to follow another user (becoming a "*follower*"), automatically getting the communication of all his/her messages. A message may be written independently, or may be in response to someone else's tweet (i.e. it is a "*reply"*). A "*retweet*" is a message promoting in the community a message of another user without altering it in any way, stressing we fully agree with it. The *hashtags* are keywords provided by the user in the tweets; a *fake user* is, usually, a humoristic duplicate of a celebrity, finally an "*influencer*" is someone who has a large number of followers (cf. Bentivegna, 2014).

A key feature of Twitter is that it is an open system, where everyone can read the tweets of other users and participate in a discussion. Many public figures, particularly politicians and showmen, have a Twitter account and anyone can write to them directly (but it is unlikely to receive a response). Therefore, Twitter is an important showcase and an inexpensive way to communicate instantly with other users of the social network, bypassing the traditional media (TV, newspapers, radio).

In the 2014 European elections, 92% of the Italian candidates had a Twitter account. In this paper, we will see how to analyze Twitter to get the sentiment towards a political figure and describe the community connected to him, although having to handle millions of tweets.

2. Political leaders on Twitter

It is interesting to note that politicians who have the highest number of Twitter followers in the world are Obama (43 million followers), followed by the Presidents of Turkey, Argentina, Colombia, Mexico, Brazil and the Queen of Jordan. The most followed politicians in Italy are Beppe Grillo (1.48 Mln followers) and Matteo Renzi (1.15 Mln followers). Then we have, by number of followers, Vendola, Bersani, Letta, Monti, Boldrini, De Magistris, Alfano. Silvio Berlusconi is not on the list because, after an initial presence on Twitter, he decided to pull out. The number of followers changes continuously, increasing or decreasing, but this number alone is not of great interest. We must consider the fact that followers are not necessarily users who share the opinions of the politician they are following and several followers could be no longer active. It is also not true that being popular on Twitter involves being popular in the country: Vendola, with 421,000 followers, should be the most popular politician following Grillo and Renzi, but this is not true.

The analyses carried out on the tweets of politicians generally use retweets, hashtags and mentions. A user who retweets a message of a politician necessarily agree with it, hence analyzing the retweets of messages we can measure the popularity of a politician. Some hashtags are of particular relevance in the political debate: *#lavoltabuona, #sfiduciamorenzie, #vinciamonoi, #vinciamopoi, #M5S* are some examples observed in the period March-May 2014 in Italy. Analyzing the popularity of hashtags can help the evaluation of political opinions; in fact, the hashtags usually can be politically labelled. The analysis of mentions of a politician is the easiest, but also coarse, way of assessing his/her popularity. Indeed, mentions and replays do not express a clear sentiment towards the politician; thus, in order to define the opinions we need to analyze the text of the tweets.

3. The information that we can get from Twitter

If we are interested in how the network judges a politician, the basic information would be the classification of the tweets as positive, negative or neutral. Of course, an expert could classify manually the tweets reading the texts, eventually discarding some tweets (ambiguous, or linked to other documents or simply jokes). If we have hundreds of thousands of tweets, this approach is clearly unfeasible and it is necessary to look for an automatic procedure or give up the classification (the last is the most common approach).

Our analysis has focused on the tweets, written between March and May 2014, which contained the name of the premier *Renzi* or the username *@matteorenzi*. We have collected, during these three months, 1,290,965 tweets, written by 136.967 users, of which 602,663 are retweets. Overall, 72% of the users wrote no more than

one message per month, while the more active users, with more than 100 messages in three months, represent only 1.5% of the users. The first large group wrote 11% of the tweets, while the small group of hyperactive users wrote as many as 43% of tweets. Each individual in a group of 56 users wrote more than 1,000 tweets in this period. This consideration should make us reflect on the difference between the sentiment of the tweets and the sentiment of the users.

Table 1 – *Most retweeted users in the period, with the number of retweets*

User	Description	Number
Matteo Renzi	Premier	64262
Matteo Salvini	Secretary of Lega Nord	8352
Gianni Kuperlo	Fake user, close to M5S	6061
La Repubblica	La Repubblica	5060
CorrieredellaSera	Corriere della Sera	5049
Andrea Scanzi	Journalist of Il Fatto Quotidiano	4987
Il Fatto Quotidiano	Il Fatto Quotidiano	4952
Francesco Manna	Blogger of Il Fatto Quotidiano	4839
Sky TG24	SKY TG 24	3893
Franco Maria Fontana	Intellectual and writer	3749
Patrizia Fiori	Quota96Scuola	3669
Fratelli d'Italia-AN	Fratelli d'Italia - Alleanza Nazionale	3622
ABATE FARIA	Blogger close to M5S	3527
Spinoza	Satirical Blog (S. Andreoli & A. Bonino)	3307
Partito Democratico	Partito Democratico	3242

What are the most retweeted users in this data? The list of the first 15 users is shown in Table 1. We also show the main hashtags and mentions in Figures 1 and 2. To make the figures readable, however, we removed "*matteorenzi*" from the hashtags, *Renzi* and *Quota96Scuola* from the mentions (*Quota96Scuola* refers to 4000 teachers who claim the right to retire). However, the number of hashtags and mentions is not very informative and of ambiguous interpretation. Therefore, when analyzing these data we should ask what the most interesting goal is.

In this paper, we identified as our main objective the understanding of the network structure of users who express opinions on a politician, identifying influencers and the relationships that bind them to each other, identifying sub-networks characterized by a particular sentiment. To achieve this goal, we must be able to classify the collected tweets with respect to the sentiment, positive or negative, on the politician. If we have more than one million of tweets, as is our case, we face a complex problem. This is the reason why all the analyses that appear in the newspapers are based on hashtags or mentions, that can be analyzed with much more ease. We must interpret the sentiment expressed by users and

influencers, analyzing the text of the tweets, also taking into account that some users, such as political parties and information agencies, are very special users.

Figure 1. *The most frequent hashtags*

Figure 2. *The most frequent mentions*

4. Mining the sentiment from 1,200,000 tweets

The data collected consist in 1,290,965 tweets written by 136.967 users from March to May 2014. For each tweet, we recorded several information about the user who wrote it and who retweeted, if any.

We created a procedure to mine the sentiment, which requires the following steps:

1. For each month, the 500 most retweeted messages were classified manually, for a total of about 1500 distinct retweets.
2. Taking into account that each of these posts had a high frequency in the archive, the first step allowed to classify 111.490 tweets of the archive. These tweets were written (more precisely retweeted) by 38.694 different users.
3. Successively, we identified all the posts, in the three months period, written from these 38.694 users, achieving 775.686 tweets.
4. We then assumed that a user approves the sentiment of the message he is retweeting (which seems obvious), and that all his posts, at least in the short term, maintain the coherence of sentiment shown in the retweet. In this way, we were able to assign the sentiments to all 775,686 tweets.
5. Eliminating some contradictory assignments, we finally got 769,982 tweets classified and 520,983 yet to be classified.

6. To manage the unclassified tweets, we built a classification model using the archive of 769,982 messages as the data sets for training and validation. The analysis can use the typical tools of text mining (cf. Applied Analytics using SAS Enterprise Miner, 2011) and a suitable classification model. In the model choice step, the policy was not to choose the model with the lowest expected classification error. Conversely, we looked for a model that was able to classify with high probability a good percentage of the data. In our data, classification trees have proven to be the most effective. In particular, we set the parameters of the tree in order to have at least 30% of the tweets with a very high probability of correct classification. To estimate the model, we used the text of the tweet and some quantitative variables that describe the user's profile.
7. The classification model, estimated in the previous step, was then applied to the remaining 520,983 messages, identifying messages with higher probability of classification (> 0.95). In this way, we were able to classify 178,243 tweets.
8. The tweets classified by the model, were written by 59,215 different users. As done in step 3, all the messages of these users were identified and classified, for a total of 377,417 tweets.
9. The tweets classified in point 8 were joined with those already classified in point 5, for a total of 1,147,399 tweets. The remaining tweets (11%) were discarded.

It is possible to make improvements that lead to change some of the above steps. It is usual, for example, that we know a priori the sentiment of some specific users (e.g. political parties or party newspapers). Another improvement consists in assigning scores to tweets or users, i.e. a non-binary value that expresses the intensity of the sentiment (e.g. an insult is a more negative sentiment than a criticism). In this way, through the evaluation of a number of scored messages, we could obtain a more reliable estimate of the user's sentiment. These corrections can lead to improve classification accuracy especially for users with many messages.

5. Analysis of the Network on Renzi's Tweets

At the end of the analyses carried out in the previous paragraph, we have available a large archive of tweets classified with respect to the sentiment. In this archive, all tweets speak, good or bad, about the premier Renzi. These data allow the analysis of the relations among users in the observed community, taking account of political opinions.

Figure 3 shows a simplified view of the network corresponding to our data. To analyze the relationships between the users, we considered only the retweets, as

they represent directed links among the users. More precisely, we considered all retweets that we were able to classify, i.e. 596,413 (97.3% of retweets), corresponding to 64,783 users. Since we could not represent a network with all of these users, we selected the most relevant nodes, showing the *influencers* and *assiduous followers*, defined in this way:

- The *influencers* are users who have been retweeted at least 700 times, during the observed period.
- An *assiduous follower* is a user who retweeted at least 25 times a specific influencer, during the observed period.

An influencer who has not assiduous followers is excluded from the graph. This is the case, for example, for "*Spinoza*" which is a satirical blog with many retweets but that has not assiduous followers and therefore does not appear on the graph.

The influencers (and his followers) who have expressed mainly positive opinions are represented by a gray square, conversely negative opinions are represented by black circles. The triangles represent the information agencies to which we have not assigned a sentiment; however, their position on the graph could be interpreted as an implicit political opinion. In the lower right, we see two sets of white diamonds: they correspond to two groups of people who are claiming certain rights and are addressing the current prime minister; the tweets in this case represent a form of pressure and do not express a clear political opinion.

The polygon size of the influencers represents the number of corresponding retweets. *Renzi* has the largest square with 64,262 retweets. The size of the followers shows how many messages they have retweeted. For all the influencers we reported their name, while the name of the followers is shown only if the number of retweets is large (>40). Being the tweets addressed to a prominent political figure, as expected, the network shows a political characterization due to the major parties and movements. In figure 3 we can easily identify sub-networks for the main political groups: *PD*, *M5S*, *Forza Italia*, *Lega Nord*, *Fratelli d'Italia*. Overall, 73% of tweets criticizes Renzi, while only 27% supports him.

The sub-network that refer to the political opinion of M5S (on the right) is broad and diversified and includes many bloggers (someone satirical), the *M5S* spokespersons and some news agencies/blogs. The *PD* sub-network is smaller, with a hierarchical structure and three fundamental references: *Matteo Renzi*, *YouDem TV* and the official account of *PD*. *Europaquotidiano* and *La Repubblica* are the closest news agencies. *Lega Nord* and *Fratelli D'Italia* are two sub-networks very active but isolated.

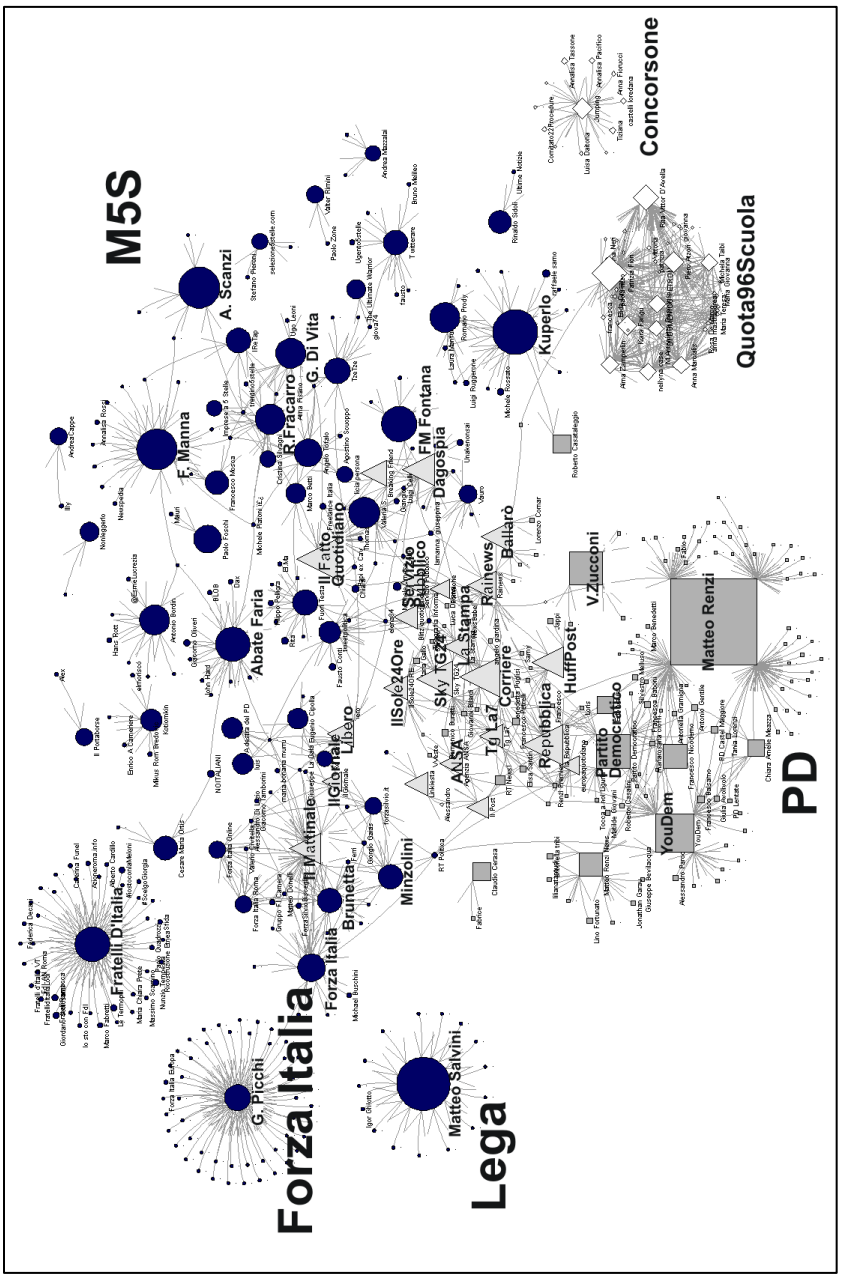


Figure 3 – *A simplified view of the full network*

The Forza Italia sub-network looks quite articulate with the presence of some very marked individualities (Picchi, Brunetta, Minzolini) and three news agencies (Il Giornale, Il Mattinale, Libero).

6. Conclusions

The paper shows how it is possible to analyze the popularity of a politician, examining millions of posts on Twitter. This was obtained using a reliable and cheap procedure, that includes text mining and statistical classification models. The information that we extracted, which also include the sentiment of users, are not typically used in the analysis of social network data. The results of our analysis also show the difference between Twitter popularity and consensus in the country.

In a deeper analysis, we can extend this approach to analyze the followers of a politician, describe the type of users in the network analyzed (always including the sentiment), and also investigate other social networks (e.g. Facebook).

Bibliographic references

SAS INSTITUTE (2011). Applied Analytics using SAS Enterprise Miner. Course notes. SAS Institute Inc., Cary, USA.

BENTIVEGNA S. (2014). La politica in 140 caratteri, Franco Angeli.

SUMMARY

Machine learning and text mining to classify tweets on a political leader

Twitter is a well-known social network. Users communicate with other users by posting short messages. These ‘tweets’ point out links among users that can be analyzed and that help to individuate “communities” who share opinions and comments. To achieve this result, we have to analyze textual data. In this paper we propose a procedure that combines machine learning techniques and text mining for the sentiment analysis on a political leader.

Agostino DI CIACCIO, Sapienza, Università di Roma, agostino.diciaccio@uniroma1.it

Giovanni Maria GIORGI, Sapienza, Università di Roma, giovanni.giorgi@uniroma1.it

Rivista Italiana di Economia Demografia e Statistica Volume LXVIII n. 3/4 Luglio-Dicembre 2014

LE ORE LAVORATE: UN'ANALISI DEI RISULTATI DELLA RILEVAZIONE SULLE FORZE LAVORO[1]

Silvia Loriga, Andrea Spizzichino

1. Introduzione

Negli ultimi anni, per diversi motivi, è cresciuto l'interesse verso la stima del numero di ore lavorate a partire dai dati della Rilevazione sulle forze di lavoro (RFL).

Un primo motivo d'interesse deriva dall'utilizzo di tali informazioni da parte della Contabilità nazionale come una misura dell'input di lavoro, regolare e non.

Inoltre, è stato recentemente costituito all'interno dell'Istat un gruppo di lavoro con l'obiettivo di sviluppare di soluzioni metodologiche per l'utilizzo integrato delle fonti statistiche per le stime dell'occupazione e delle ore lavorate; le fonti statistiche a cui ci si riferisce in questo contesto sono i dati derivanti da indagini Istat sulle famiglie, tra cui principalmente la Rilevazione sulle forze di lavoro, sulle imprese e da archivi amministrativi.

Infine, è stata recentemente costituita una Task Force Eurostat su 'Measurement of Absences and Working time in the EU LFS' con l'obiettivo di migliorare la qualità e la comparabilità di tali variabili a livello europeo.

Il presente lavoro si colloca in questo contesto, con l'obiettivo di valutare la qualità di queste variabili rilevate nella RFL; in particolare, viene riportata prima una descrizione delle domande relative alle ore lavorate nel questionario RFL; vengono poi descritti alcuni miglioramenti ottenuti inserendo dei *warning* nel questionario volti a ricordare all'intervistato la presenza di festività nella settimana di riferimento; infine vengono effettuate alcune elaborazioni, cercando di riprodurre le analisi presentate da Francia e Germania in occasione di un workshop dedicato alla LFS nel 2012[2].

[1] Il paper è frutto di un lavoro di elaborazione e sintesi congiunto tra gli autori, tuttavia i paragrafi 1,2 e 3 sono da attribuibile a Silvia Loriga, i par.4,5 e 6 ad Andrea Spizzichino.

[2] 7th Workshop on Labour Force Survey Methodology svoltosi a Madrid il 10 e 11 maggio 2012

2. I quesiti sulle ore lavorate

Nel questionario della Rilevazione sulle forze di lavoro, la parte dedicata alla definizione del numero di ore lavorate viene dopo il modulo relativo alla tipologia d'orario lavorativo (tempo pieno/ tempo parziale). La prima domanda è sul numero di ore lavorate abitualmente, al netto dell'interruzione per i pasti e gli spostamenti casa-lavoro; per coloro che non sanno rispondere a tale quesito o dichiarano di avere un orario molto variabile viene chiesto, con riferimento alle ultime 4 settimane, il numero di ore lavorate mediamente a settimana.

Per la definizione delle ore effettivamente lavorate nella settimana di riferimento si scende più nel dettaglio: prima di chiederne il numero si domanda se in quella settimana ha lavorato meno ore o più ore rispetto alle ore che lavora di solito e il motivo prevalente per cui ha lavorato di più o di meno; viene poi chiesto all'intervistato se ha svolto ore di straordinario retribuito e/o non retribuito e il numero delle une e delle altre; solo infine si chiede il numero di ore effettivamente lavorate nella settimana di riferimento.

Questo insieme di domande consente di approfondire l'analisi sulle ore lavorate, andando inizialmente a distinguere il numero di ore lavorate abitualmente da quelle lavorate effettivamente nella settimana di riferimento e poi ad identificare le motivazioni e le intensità delle differenze.

3. Ore non lavorate per giorni di festività e di ferie

Nella figura 1 vengono riportati, per le 52 settimane del 2010, il numero di occupati, le ore effettivamente lavorate (moltiplicate per 10 per questioni grafiche), il numero di occupati che hanno lavorato meno rispetto all'orario abituale e, tra questi, coloro che hanno lavorato meno per festività o ferie.

Emerge che la rilevazione riesce a cogliere la stagionalità del fenomeno e soprattutto i picchi di assenze dovute alle ferie estive e alle festività che si presentano nell'arco dell'anno.

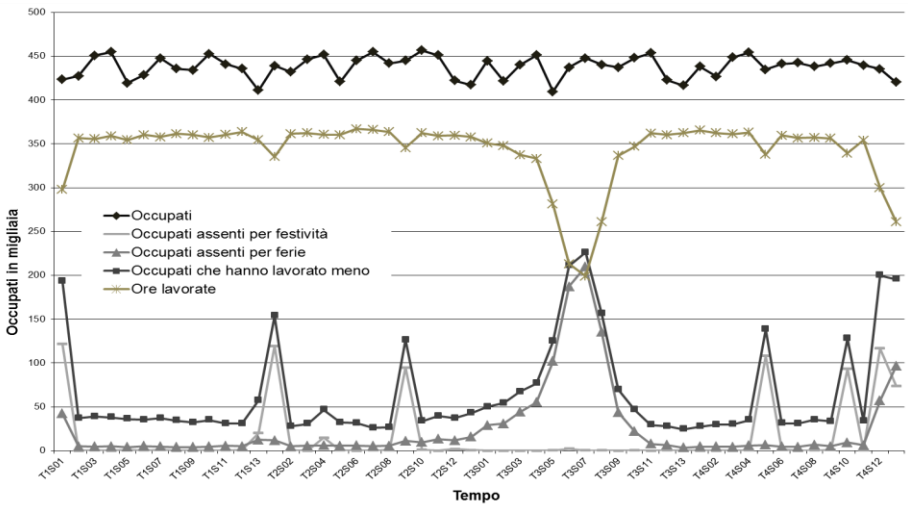
In corrispondenza delle settimane in cui si presentano festività o in cui è più diffuso l'utilizzo di ferie, il numero medio di ore effettivamente lavorate è sempre più basso e la quota di occupati che dichiarano di aver lavorato meno per uno di questi motivi è sempre più alta.

Nel corso dell'anno i picchi vengono raggiunti in estate, in particolare nelle 2 settimane centrali di agosto, quando quasi la metà degli occupati dichiara di aver lavorato meno ore per ferie e in diverse settimane dell'anno, in particolare quelle in cui cade il Natale e l'Epifania, in cui oltre un terzo degli occupati dichiara di aver lavorato meno ore per la presenza di un giorno di festa nella settimana di riferimento.

Nonostante la dinamica dell'andamento settimanale del numero di ore lavorate e delle assenze dal lavoro sia coerente con le attese, ci si aspettava una maggiore intensità nel fenomeno, in particolare per determinati gruppi di occupati in alcune settimane. Dall'analisi delle dichiarazioni di assenza dal lavoro per festività o ferie, rispetto alle caratteristiche sia dell'intervista sia dell'intervistato, risulta che:

- Non emergono particolari differenze all'aumentare della distanza tra la settimana dell'intervista e quella di riferimento; sembrerebbe dunque non esserci un effetto "memoria".
- Le interviste in CAPI rilevano una minore quota di assenze per ferie rispetto a quelle in CATI (è opportuno osservare che ciò dipende dalla maggiore incidenza delle interviste telefoniche nei mesi estivi).
- Non emergono differenze tra interviste proxi e dirette.
- Per i principali macro-settori d'attività le differenze sono coerenti con le attese, con gli occupati nei settori secondario e terziario che beneficiano più degli occupati in agricoltura dei giorni di festa.

Figura 1 – *Ore lavorate e occupati nel complesso e per motivo d'assenza nelle 52 settimane del 2010.*



- Le differenze tra settori all'interno del terziario sembrano non riflettere a pieno le attese. Sono stati confrontati per esempio gli occupati in hotel e ristoranti, che nel terziario sono quelli che godono meno dei giorni di festività, e quelli in banche e assicurazioni, che sono quelli che ne godono di più; nelle settimane in cui, presumibilmente, tutti gli occupati in banche e assicurazioni dovrebbero aver

beneficiato di un giorno di festività, le differenze tra i due settori non sono troppo marcate e tra i dipendenti di banche e assicurazioni al massimo il 35% dichiara di aver beneficiato di festività.

L'impressione che ci possa essere una sottostima degli eventi di assenza per ferie e in particolare festività con conseguente sovrastima del numero di ore effettivamente lavorate nel complesso, ha spinto a introdurre delle modifiche nella sezione di questionario in cui si rileva il numero di ore effettivamente lavorate.

4. Miglioramenti nell'informazione sulle ore non lavorate per festività

In occasione di una indagine pilota, condotta su 500 famiglie, svolta con riferimento alla prima settimana di novembre 2012, si è pensato di apportare delle modifiche alle domande del questionario sulle ore lavorate; in particolare si è cercato di aiutare l'intervistato a ricordare se aveva goduto di festività o ferie segnalando la presenza di una festività nella settimana di riferimento. La frase del questionario che è stata modificata è la seguente (in grassetto la modifica che è stata introdotta):

Le prossime domande si riferiscono alle ore di lavoro svolte "LA SCORSA SETTIMANA", cioè la settimana che va "DA LUNEDI' … A DOMENICA…". ***Consideri che "LA SCORSA SETTIMANA" c'è stata la festività del primo novembre*** *e tenga conto anche di eventuali ferie, malattia, straordinario, etc.*

I risultati sono stati soddisfacenti al punto che dal I trimestre del 2013 è stato inserito correntemente il *warning* in presenza di festività nella settimana di riferimento.

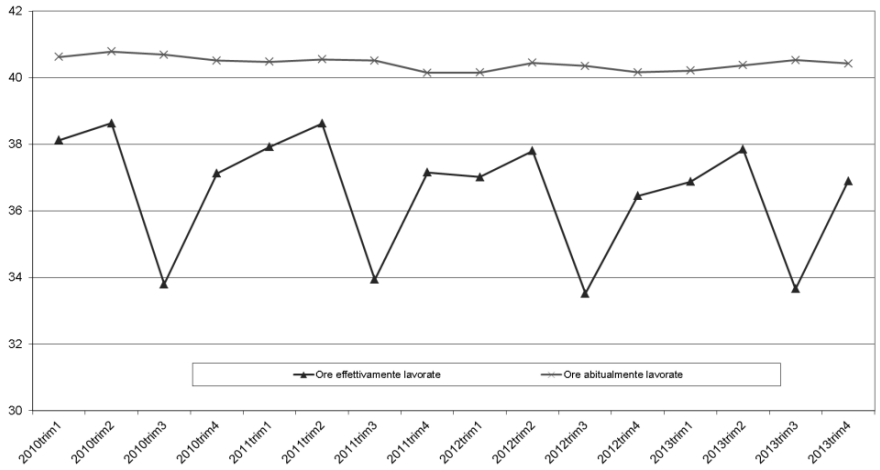
Nell'analizzare le serie storiche sul numero ore effettivamente lavorate bisogna tener presente le difficoltà derivanti dalla distribuzione delle festività nei giorni dell'anno; succede infatti che:

- La stessa festività sia presente a distanza di un anno in trimestri diversi (Es. la Pasqua).
- In un anno ci siano due giorni di festa nella stessa settimana di riferimento e nel successivo no (Es. Santo Stefano e primo dell'anno).
- I giorni di festa cadano in un anno di sabato o domenica per cui non vengono considerati.

Tenendo a mente queste considerazioni, la figura 2 mostra l'andamento del numero di ore lavorate abitualmente ed effettivamente nella settimana di

riferimento dagli occupati full-time[3] tra il 2010 e il 2013; si può notare che a fronte di una sostanziale stabilità del numero di ore abituali si registra una diminuzione di quelle effettive a partire dalla fine del 2012 e per tutto il 2013.

Figura 2 – *Ore lavorate abitualmente ed effettivamente dagli occupati full-time tra il 2010 e il 2013.*



In media annua, sempre tra gli occupati full-time, la differenza tra ore abitualmente ed effettivamente lavorate passa da 3,7 nel 2010 a 4,1 nel 2013 e la percentuale di occupati che dichiara assenze per festività passa dal 3,6% al 4,2%, a dimostrazione di quanto, grazie all'introduzione del *warning*, venga colto meglio questo tipo di assenza.

5. Stima dei giorni di ferie

Dopo aver analizzato le assenze per festività si è cercato di stimare il numero medio di giorni di ferie goduti nell'arco dell'anno, replicando le analisi presentate da Francia e Germania in occasione del 'Workshop on Labour Force Survey Methodology' del 2012.

Per i soli occupati a tempo pieno, che dichiarano di aver lavorato nella settimana di riferimento meno del solito a causa delle ferie, viene definito il totale di ore settimanali non lavorate per ferie, sottraendo alle ore abituali le ore effettive e aggiungendo le ore di straordinario.

[3] Vengono considerati i soli full-time e non gli occupati nel complesso per depurare il numero totale di ore lavorate dalla crescita dei contratti part-time che è caratteristica degli anni presi in considerazione.

Questo totale viene diviso per il numero di occupati full-time e moltiplicato per il numero di settimane in un anno (52) ottenendo il numero di ore non lavorate per ferie dal un occupato full-time durante l'anno. Dividendo per il numero medio di ore lavorate quotidianamente, che viene stabilito pari a 8, si ottiene una stima dei giorni non lavorati per ferie che per il 2010 e 2013 risultano pari a 12,5 e 12,8 (Tabella 1).

Tabella 1 – *Stima del numero di giorni non lavorati per ferie. Anni 2010 e 2013*.

Anno	Ore settimanali non lavorate per ferie nel complesso	Occupati full-time	Numero settimane	Ore non lavorate per ferie per occupato full-time	Giorni non lavorati per ferie per occupato full-time
2010	37171	19263	52	100.3	12.5
2013	35616	18141	52	102.1	12.8

Approfondendo ulteriormente l'analisi si è notato che alcuni occupati che dichiarano di essere stati assenti per festività, dichiarano un numero di ore non lavorate nella settimana di riferimento che presumibilmente supera una giornata lavorativa (>10) anche se nella settimana era presente solo un giorno di festività; ciò deriva probabilmente dall'unione di giorni di ferie con giorni di festa (effetto 'ponte')[4]. Queste ulteriori ore di ferie, stimate sottraendo in modo cautelativo 10 ore al totale di ore non lavorate nella settimana di riferimento, sono state aggiunte a quelle considerate precedentemente e hanno determinato altri 0,5 giorni di ferie annuali per il 2010 e 1 giorno per il 2013, che sommati ai giorni di ferie già stimati danno in tutto 13 e 13,8 giorni di ferie nei 2 anni considerati.

I risultati evidenziano chiaramente come a distanza di 4 anni sia rimasto sostanzialmente invariato il numero di giorni di ferie desumibili direttamente (da 12,5 a 12,8) ma siano aumentati quelli che si ottengono indirettamente (da 0,5 a 1) grazie al miglioramento dell'informazione relativa a chi beneficia di giorni di festa nella settimana di riferimento.

[4] Nella sezione di questionario in cui viene chiesta la causa dell'assenza dal lavoro si fa riferimento solo al motivo principale dell'assenza.

Tabella 2 – *Stima del numero di ulteriori giorni non lavorati per ferie. Anni 2010 e 2013.*

Anno	Ore settimanali non lavorate per ferie nel complesso	Occupati full-time	Numero settimane	Ore non lavorate per ferie per occupato	Giorni non lavorati per ferie per occupato
2010	1547	19263	52	4.2	0.5
2013	2754	18141	52	7.9	1.0

6. Conclusioni

In questo lavoro viene descritto uno studio condotto in Istat con riferimento alla stima delle ore lavorate a partire dai dati della Rilevazione sulle forze di lavoro.

Di particolare interesse sono i miglioramenti nella qualità delle informazioni raccolte ottenuti grazie all'inserimento di un *warning* volto a ricordare all'intervistato la presenza di un giorno di festa nella settimana di riferimento; va sottolineato come con costi praticamente nulli si sia ottenuto un significativo miglioramento dell'informazione sulle ore effettivamente lavorate.

L'interesse per le variabili sulle ore lavorate non si limita alla sola rilevazione italiana sulle forze di lavoro ma è attualmente oggetto di studio da parte di Eurostat che ha costituito una Task Force dedicata a 'Measurement of Absences and Working time in the EU LFS'. Nell'ottica di migliorare la qualità e la comparabilità delle variabili sulle ore di lavoro il *warning* introdotto nel 2013 nel questionario italiano dovrebbe diventare pratica diffusa a livello europeo.

Quanto finora prodotto non vuole rappresentare un punto d'arrivo bensì uno stimolo ad approfondire ulteriormente l'analisi e la ricerca su questa tematica che raccoglie sempre più interesse; in quest'ottica ulteriori sviluppi sono in programma in termini di analisi e di confronto delle informazioni raccolte dalla RFL con le informazioni desunte dai dati di fonte amministrativa.

Riferimenti bibliografici

KORNER T, WOLFF L. 2012. Quality issues regarding the number of hours actually worked in the French and German LFS, 7th European WS on LFS Methodology, Madrid, 10-11 May 2012.

ISTAT. 2006. *La rilevazione sulle forze di lavoro: contenuti, metodologie, organizzazione*. (Metodi e norme, n. 32).

ISTAT. 2011. I tempi del lavoro. Collana Argomenti n°40, Istat.

SUMMARY

Worked hours: an analysis of the results of the Italian Labour Force Survey

In recent years, in Istat, the interest for the number of worked hours estimated by the Labour Force Survey (RFL) increased a lot, due to several reasons: because of the role working hours play in the debate on labor productivity and wages, because of the relevance of the time dedicated to work in the organization of individual and family life and, finally, because the number of worked hours is used as an indicator of labour input alongside the number of employed individuals in the National Accounts framework.

Recently, reproducing studies made by other National Statistical Institutes, we analyzed the various information on working time available in the RFL, especially focusing on the hypothesis that the survey overestimates the number of hours actually worked due to some kind of memory effect resulting in an under-reporting of absences.

Taking advantage of a pilot survey conducted in the first week of November 2012 and currently since January 2013, changes were introduced in the section of the questionnaire concerning working hours, trying to improve the memory of the respondent on worked hours and absences during the reference week.

The results are satisfactory and the quality of information on worked hours and absences has improved.

In this paper we report the results of the analysis on actual worked hours estimated by the RFL and on the improvements achieved thanks to the changes made to the questionnaire.

Silvia LORIGA, Istat, siloriga@istat.it
Andrea SPIZZICHINO, Istat, spizzich@istat.it

OCCUPAZIONE FEMMINILE: L'OLANDA UN ESEMPIO VIRTUOSO PER L'ITALIA?[1]

Rosa Calamo, Thaís García Pereiro

1. Introduzione

L'occupazione femminile in Italia resta ancora al di sotto dei parametri quantitativi stabiliti dal Consiglio di Lisbona del 2000: un tasso di occupazione medio europeo superiore al 60% da raggiungersi entro il 2010. Le ragioni di questo ritardo sono molteplici. Tra tutte si vuole focalizzare l'attenzione sulla difficoltà di conciliazione dovuta alla carenza di servizi per l'infanzia.

Questo lavoro nasce dalle questioni sollevate da un'intervista rilasciata di recente dal direttore generale del Fondo Monetario Internazionale che ha bocciato l'Italia sul tema dell'occupazione femminile indicando l'Olanda quale paese modello dell'Unione Europea.

L'obiettivo centrale di questo articolo è quello di analizzare l'occupazione femminile in entrambi i paesi e il suo andamento nel tempo, concentrando l'attenzione sul peso del lavoro part time come alternativa per la conciliazione famiglia-lavoro e sulla disponibilità di centri per la cura dell'infanzia quale supporto alla continuità dell'impiego.

I dati utilizzati sono aggregati e secondari, costituiti dalle serie storiche:

Italia, Olanda, EU15: Eurostat (datawarehouse).

- *Tasso di Occupazione Femminile (TOF), per classi di età e livello educativo;*
- *Percentuale di Lavoro Part-Time Femminile (%PTF) e per classi di età;*
- *Percentuale di Part-Time Femminile Involontario (%PTFI).*

Italia: Istituto Nazionale di Statistica ISTAT, Istat (datawarehouse).

Olanda: Eurostat - Feasibility study on the availability of comparable child care statistics in the EU.

[1] L'esecuzione complessiva del lavoro va intesa svolta dagli autori (citati in ordine alfabetico) in stretta collaborazione: tuttavia, si attribuiscono i par.1 e 3 a R. Calamo, i par.2 e 4 a T. García Pereiro, il par.5 va attribuito a entrambi gli autori.

2. Letteratura

Il lavoro part-time è stato considerato nei Paesi Bassi come il modo ideale per conciliare lavoro ed assistenza in casa. In linea con questa visione, il governo ha introdotto una legislazione che garantisce gli stessi diritti ai lavoratori a tempo pieno e parziale (Jaumotte, 2003).

Sono soprattutto le donne che sfruttano la possibilità di lavorare part-time ed è per questo motivo che il tasso di occupazione femminile nei Paesi Bassi e tra i più alti dell'Europa (Schettkat e Yocarini, 2001).

La crescita del tasso di occupazione e l'alta percentuale di posti di lavoro part-time è strettamente collegata con la percentuale di madri che lavorano. La maggior parte delle donne nei Paesi Bassi, infatti, continua a lavorare anche dopo la nascita dei loro figli (Kalwij, 2000).

In Italia negli ultimi anni si è registrato un aumento del numero di occupati e ciò è dovuto proprio all'aumento dell'occupazione femminile (Saraceno, 2003). Il tasso di occupazione femminile in Italia è aumentato drasticamente negli ultimi 20 anni, guadagnando dieci punti percentuali ma non ha raggiunto gli obiettivi previsti dalla strategia di Lisbona ed è ancora lontano dalla media europea (UE - 28: 58,5 % nel 2011).

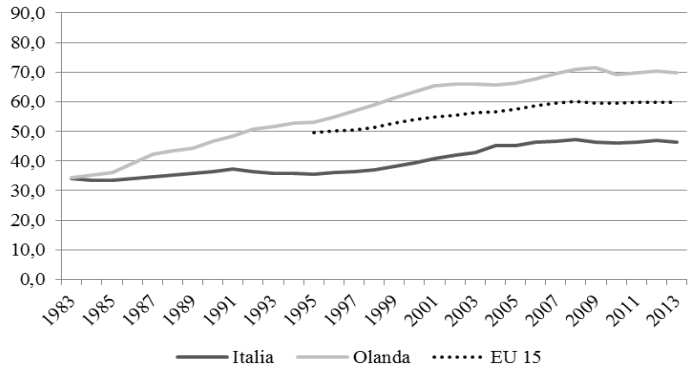
Come evidenziato dal lavoro di Moreno Minguez (2008) in Italia la decisione di partecipare al mercato del lavoro dipende in gran parte dai carichi familiari e soprattutto dal livello educativo: più è alto, più è probabile che le donne siano occupate.

Una caratteristica peculiare del mercato del lavoro femminile in Italia è quella di presentare forti differenze territoriali (le più alte d'Europa). C'è chi attribuisce il basso tasso di occupazione femminile in Italia proprio alla scarsa partecipazione al mercato del lavoro nel Mezzogiorno (Scherer e Reyneri, 2008).

3. Occupazione

Dal 1986 il Tasso di Occupazione Maschile (TOM) in Italia risulta più basso che in Olanda e in EU-15 e a partire dal 2008 è sceso al di sotto della soglia del 70% mentre in Olanda si trova al di sopra del 80% sin dall'anno 2000. Come si rileva dalla Figura 1 il Tasso di Occupazione Femminile (TOF) aumenta considerevolmente dall'inizio degli anni ottanta. L'incremento è maggiore in Olanda, dove esso passa da 34% nel 1983 al 70% nel 2013. In Italia l'incremento è di soli 12,5 punti percentuali in trent'anni.

Figura 1 – *Tasso di Occupazione Femminile (TOF). Italia, Olanda, EU15. Anni 1983-2013.*



Fonte: Eurostat.

Per quanto riguarda il Tasso di Occupazione Femminile per gruppi di età (Tabella 1) si rileva che fino al 1983 per la fascia di età 30-34 anni era più alto in Italia (49,8% vs. 36,6%). A partire da questa data quello dell'Olanda cresce significativamente (si raddoppia) attestandosi nel 2013 sull'80%. L'incremento in Italia non raggiunge 10 punti percentuali. Per la fascia di età 35-39 anni fino al 1983 era più alto in Italia (44,4% vs. 39,1%). A partire da questa data il TOF dell'Olanda cresce significativamente (si raddoppia) attestandosi nel 2013 sull'80%. L'incremento in Italia raggiunge 18 punti percentuali. Per la fascia di età 40-59 anni fino al 1983 il TOF era simile in entrambi i Paesi. In Olanda cresce di 44 punti percentuali. Mentre in Italia l'incremento raggiunge 25 punti percentuali. La differenza tra la crescita dei TOF di questo gruppo di età è minore rispetto ai gruppi più giovani.

Tabella 1 – *Occupazione femminile per gruppi di età (30-59 anni). Italia, Olanda, EU15. Anni 1983-2013.*

		1983	1993	2003	2013
EU15	30-34	-	-	68,6	70,2
	35-39	-	-	69,7	71,8
	40-59	-	-	62,7	69,7
Italia	30-34	49,8	51,4	57,8	57,3
	35-39	44,4	52,8	59,5	62,2
	40-59	31,7	36,7	47,1	56,9
Olanda	30-34	36,6	61,1	78,5	79,8
	35-39	39,1	57,8	74,3	79,8
	40-59	29,8	47,2	64,5	73,6

Fonte: Eurostat.

L'evoluzione del TOF fino al primo livello d'istruzione secondaria è pressoché stabile per l'Italia e per l'EU15. Mentre in Olanda aumenta di ben 10 punti percentuali. Per quanto riguarda, invece il secondo livello d'istruzione secondaria esso tende a crescere in tutti i Paesi ma in Olanda è più alto e aumenta maggiormente rispetto all'Italia. Il TOF delle donne con istruzione universitaria risulta il più alto in riferimento al livello d'istruzione in entrambi i Paesi, sempre al di sopra del 70%. In Italia esso si riduce di 2 punti percentuali nel 2013, mentre in Olanda cresce fino all'86% (9,5 punti percentuali).

Tabella 2 – *Tasso di Occupazione Femminile per livello di istruzione. Italia, Olanda, EU15. Anni 1996-2000-2004-2008-2013.*

		1996	2000	2004	2008	2013
EU15	Fino al I livello istruzione secondaria	37,2	40,2	40,2	41,0	38,5
	II livello istruzione secondaria	59,0	62,1	63,4	65,9	64,3
	Istruzione universitaria	75,9	77,9	78,9	80,0	78,4
Italia	Fino al I livello istruzione secondaria	25,8	26,5	29,5	29,6	29,1
	II livello istruzione secondaria	50,3	53,5	57,6	58,6	54,1
	Istruzione universitaria	74,0	74,2	76,3	73,8	71,6
Olanda	Fino al I livello istruzione secondaria	38,7	46,8	48,3	53,0	50,9
	II livello istruzione secondaria	65,0	71,6	71,6	76,2	73,0
	Istruzione universitaria	76,3	81,4	82,4	85,5	85,8

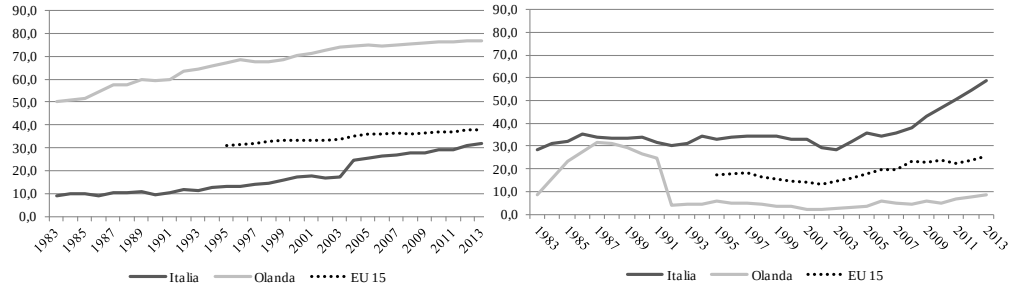
Fonte: Eurostat.

4. Part time femminile

Il part-time è diventato un modo di lavorare, e di vivere, sempre più diffuso. Anche perché riesce, più di altri strumenti, a coinvolgere nel mondo dell'occupazione chi, come le donne, rischia altrimenti di rimanerne fuori. In Europa, oggi, un lavoratore su cinque ha un contratto a tempo parziale. In alcuni paesi la percentuale è anche maggiore. Coinvolge soprattutto la componente femminile ma in questi ultimi tempi comincia a interessare anche figure manageriali e di responsabilità.

Il gap del Part Time Femminile (PTF) (Figura 2) tra i due Paesi rimane costante nel tempo: era di 41 punti percentuali nel 1983 e raggiunge i 45 punti trent'anni dopo. Mentre in Olanda la percentuale di donne che lavorano part-time nel 2013 è 77% in Italia è solo 31% (e occorre sottolineare che tale incremento riguarda proprio gli ultimi 10 anni).

Figura 2 e 3 – *Percentuale di Part-Time Femminile (PTF) e Part-Time Femminile Involontario (15-64 anni). Italia, Olanda, EU15. Anni 1983-2013.*



Fonte: Eurostat.

L'aumento del part time è associato ad una crescita del part time "involontario", identificato con quanti dichiarano di svolgere un lavoro a tempo parziale in mancanza di occasioni lavorative a tempo pieno

Il part-time involontario (Figura 3) è prevalente in Italia e registra una crescita di ben 30 punti percentuali negli ultimi 8 anni raggiungendo il 60%. In Olanda si attesta intorno al 10%.

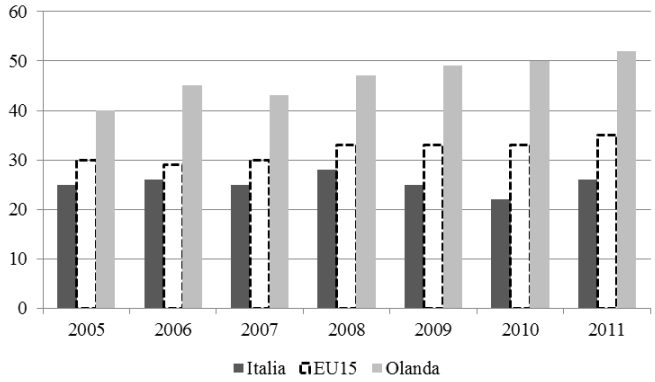
5. Asili nido

L'importanza di fornire adeguati servizi per l'infanzia è stata riconosciuta a livello Europeo, infatti l'Agenda di Lisbona ha definito alcuni obiettivi espliciti riguardo la loro fornitura: coprire, entro il 2010, almeno il 90% dei bambini fra 3 e 6 anni, ed almeno il 33% dei bambini sotto i 3 anni.

In realtà, la diffusione di tali servizi differisce in modo notevole all'interno degli Stati membri ed in molti Paesi (tra cui l'Italia) si è ancora molto lontani dall'obiettivo fissato.

Come evidenziato dalla Figura 4 la percentuale di bambini di meno di 3 anni che usufruiscono di asili nido cresce in maniera continua in Olanda passando da 40% a 52% mentre in Italia l'andamento è solo di leggera e discontinua crescita passando da 25 a 27%.

Figura 4 – *Percentuale di bambini di meno di 3 anni che usufruisco di asili nido. Italia, Olanda, EU15. Anni 2005-2011.*



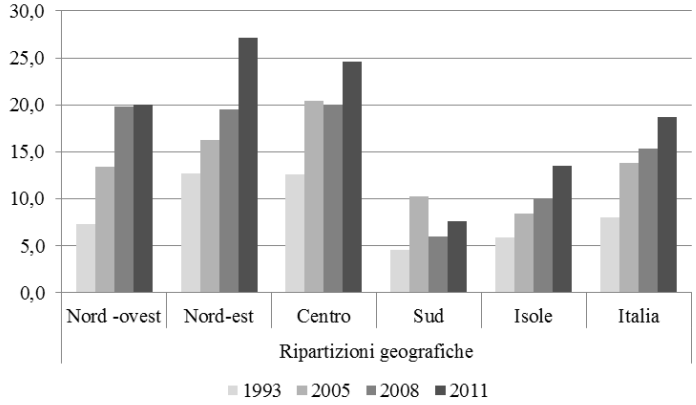
Fonte: Eurostat.

In Olanda il numero di bambini che usufruisce di asili nido parte da 40 per mille nel 1990 e arriva a 100 nel 1996. La percentuale di primi figli che hanno frequentato l'asilo nei primi 4 anni di età si è quasi raddoppiata nelle coorti di nascita considerate.

In Italia, secondo i dati forniti dall'Istat, nonostante il generale ampliamento dell'offerta pubblica, la quota di domanda soddisfatta è ancora limitata rispetto al potenziale bacino di utenza: gli utenti degli asili nido sono passati dal 8,0% dei residenti tra zero e due anni dell'anno scolastico 1993 al 18% del 2011. Sono evidenti notevoli differenze territoriali (Figura 5). Sebbene la tendenza è all'aumento, il Nord-est risalta per il suo sviluppo (da 12% a 27% in 18 anni) ed il Sud per la sua stazionarietà (da 4,5% a 7,6%).

La percentuale di iscritti alla scuola dell'infanzia per tipo di scuola e ripartizione geografica dell'anno scolastico 2010/2011 (Istat) evidenzia il maggiore peso degli asili privati nel Nord-Ovest (38,5%) e Nord-Est (46,7%), mentre nel Mezzogiorno prevale l'offerta pubblica (77,0% nel Sud e 80,6% nelle Isole).

Figura 5 – *Bambini 0-2 anni iscritti all'asilo nido per ripartizione geografica (per 100 bambini di 0-2 anni). Anni 1993, 2005, 2008 e 2011.*



Fonte: Istat.

Conclusioni

Il TOF è cresciuto in entrambi i paesi ma l'incremento è maggiore in Olanda (da 34% nel 1983 a 70% nel 2013) che in Italia (da 34% nel 1983 a 46,5% nel 2013).

Il TOF delle donne con istruzione universitaria risulta il più alto in riferimento al livello d'istruzione in entrambi i Paesi, sempre al di sopra del 70%.

Il gap del PTF tra i due Paesi è di ben 40 punti con prevalenza del part-time involontario in Italia.

La percentuale di bambini con meno di 3 anni che usufruiscono di asili nido in Olanda è 52% mentre in Italia è solo 27% con grosse differenze territoriali (al Sud si raggiunge appena il 7,6% e si tratta di scuole pubbliche).

Resta evidente la difficoltà di conciliazione lavoro-famiglia in Italia con un livello di part-time così basso e una presenza di asili nido insoddisfacente ciò è vero soprattutto per le donne del Sud, che hanno a che fare con un mercato del lavoro più difficile e un sistema quasi inesistente per la custodia dei bambini.

Lo studio della situazione dell'Olanda ci permette di concludere che un welfare ed un mercato del lavoro più favorevole consentono alle donne una maggiore occupazione ma cruciale è il ruolo che deve svolgere lo Stato in relazione alle misure pubbliche di sostegno. C'è ora in Italia una nuova legge delega dell'attuale governo che affronta anche il problema della occupazione femminile. Bisognerà valutare l'adeguatezza di queste misure ad un Paese così eterogeneo.

Riferimenti bibliografici

JAUMOTTE, F. 2003. Female Labour Force Participation: Past Trends and Main Determinants in OECD Countries. *OECD Economics Department Working Papers* 376.

KALWIJ, A.S. 2000. The Effects of Female Employment Status on the Presence and Number of Children. *Journal of Population Economics*, 13, pp. 221-239.

MORENO MÍNGUEZ, A. 2008. El reducido empleo femenino en los estados del bienestar del sur de europa Un análisis comparado. *Revista Internacional de Sociología (RIS,)* VOL. LXVI, Nº 50, pp. 129-162.

SCHERER, S., REYNERI, E. 2008. Come è cresciuta l'occupazione femminile in Italia: fattori strutturali e culturali a confronto. *Stato e Mercato*, 2, pp. 183-216.

SARACENO, C. 2003. La conciliazione di responsabilità familiari e attività lavorative in Italia: paradossi ed equilibri imperfetti. *Polis*, 17(2), pp. 199-228.

SCHETTKAT, R., YOCARINI, L. (2001). Education Driving the Rise in Dutch Female Employment: Explanations for the Increase in Part-time Work and Female Employment in the Netherlands, Contrasted with Germany. *IZA Discussion Papers*, 407, Institute for the Study of Labor (IZA).

SUMMARY

Female employment: Netherlands a positive example for Italy?

Women's employment rates in Italy are still below the quantitative parameters set by the Lisbon Council in 2000: an average employment rate higher than 60% to be achieved by 2010. The reasons for this delay are manifold, specially important is the difficulty of reconciling work and family life. Recently, the General Director of the International Monetary Fund pointed out Netherlands as a model country for Italy on the subject of women's employment. Thus, the main purpose of this article is to analyze and compare women's employment in both Italy and Netherlands, focusing on the evolution of employment rates, the role of part-time jobs as an alternative for reconciling work and family and the availability child-care centers as a support to employment continuity.

Rosa CALAMO, Università degli Studi di Bari, rosa.calamo@uniba.it
Thaís GARCÍA PEREIRO, Università degli Studi di Bari, tgarciapereiro@libero.it

A COMPOSITE INDEX FOR MEASURING ITALIAN REGIONS' DEVELOPMENT OVER TIME[1]

Matteo Mazziotta, Adriano Pareto

1. Introduction

In the last years, many have been the attempts of public and private institutions as well as individual researchers to construct composite indices of development, well-being or societal progress (Bandura, 2008).

The most famous cross-national index of socio-economic well-being is probably the *Human Development Index* (HDI), developed by the United Nations (UNDP, 2010). In Italy, important examples of regional indices of well-being are the *Index of Regional Quality of Development* "QUARS", proposed by the campaign "Sbilanciamoci!" (Gnesi *et al.*, 2010), and the *Quality of Life Index*, published by the economic newspaper "Il Sole 24 Ore" (Lun *et al.*, 2006). Furthermore, since 2010, the National Institute of Statistics (Istat) and the National Council for Economy and Labour (Cnel) have launched a project, called BES (Benessere Equo Sostenibile), for constructing a set of indicators of the various dimensions of well-being in the Italian regions (Istat, 2014).

In general, a composite index is based on several indicators or sub-indices (pillars). These indicators or sub-indices are aggregated by analytical methods to give an overall score for each country or geographical area. The results are used to either create a ranking or to simply summarize the data (OECD, 2008).

An important issue concerning the composite index construction is the level of comparability of the data across units and over time (Tarantola, 2008). Comparability of the values of a composite index firstly depends on the normalization rule. All the normalization methods allow for space comparisons, whereas time comparisons may be difficult to make or to interpret.

Comparisons over time may be absolute or relative. We say that a time comparison is 'relative' when the composite index values, at time *t*, depend on one or more endogenous parameters (e.g., mean and variance of the individual

[1] The paper is the result of combined work of the authors: M. Mazziotta has written Sects. 3 and 4; A. Pareto has written Sects. 1 and 2.

indicators at time *t*). Similarly, we say that a time comparison is 'absolute' when the composite index values, at time *t*, depend on one or more exogenous parameters (e.g., minimum and maximum of the individual indicators fixed by the researcher).

The QUARS and the "Il Sole 24 Ore" *Quality of Life Index* allow only for relative comparisons since they are based exclusively on values of the individual indicators for the year of reference. In the case of the HDI, the problem has been overcome by using a re-scaling of the indicators in the range (0; 1) with limits independent from the observed values in a given year. This solution may lead to future values outside the range and the only alternative is to recalculate the index values for the past years (Tarantola, 2008).

In this paper, we present a variant of the Mazziotta-Pareto Index (De Muro *et al.*, 2010) that allows to make absolute comparisons over time, in a not full compensatory perspective (OECD, 2008). In Sections 2 a brief description of the method is reported; in Section 3 an application to indicators of socio-environmental development in Italy is presented.

2. The adjusted MPI

The Mazziotta-Pareto Index (MPI) is a non-compensatory[2] composite index based on a standardization of the individual indicators, at the reference time, that makes the indicators independent of the variability (normalized indicators have a mean of 100 and a standard deviation of 10). Therefore, all the individual indicators are assigned equal weights, but only relative time comparisons (with respect to the mean) are allowed (De Muro *et al.*, 2010).

In order to appreciate absolute changes over time, we propose a different procedure of normalization of data based on a re-scaling of the individual indicators according to two 'goalposts', i.e., a minimum and a maximum value which represent the possible range of each indicator for all time periods considered (Mazziotta and Pareto, 2013b).

The steps for computing the variant of MPI for time comparisons, namely Adjusted MPI (AMPI), are given below.

[2] A non-compensatory composite index is a index based on the assumption of 'non-substitutability' of the individual indicators or pillars, i.e., they have all the same 'importance' and a full compensation among them is not allowed. In a non-compensatory approach, all the dimensions of the phenomenon must be balanced and an aggregation function that takes unbalance into account, in terms of penalization, is often used. A compensatory approach involves the use of additive methods, such as the arithmetic mean. A non-compensatory approach generally requires non-linear methods, such as the geometric mean or the Multi-Criteria Analysis (OECD, 2008).

Given the matrix $\mathbf{X}=\{x_{ij}\}$, we calculate the normalized matrix $\mathbf{R}=\{r_{ij}\}$ as follow:

$$r_{ij} = \frac{(x_{ij} - \text{Min}_{x_j})}{(\text{Max}_{x_j} - \text{Min}_{x_j})} 60 + 70 \qquad (1)$$

where Min_{x_j} and Max_{x_j} are the 'goalposts' for the indicator *j*. If the indicator *j* has negative 'polarity'[3], the complement of (1) with respect to 200 is calculated. In both cases, the range of the normalized values is (70; 130).

Denoting with M_{r_i} and S_{r_i}, respectively, the mean and the standard deviation of the normalized values of the unit *i*, the generalized form of the AMPI is given by:

$$\text{AMPI}_i^{+/-} = \text{M}_{r_i} \pm \text{S}_{r_i}\text{cv}_i$$

where $\text{cv}_i = \text{S}_{r_i} / \text{M}_{r_i}$ is the coefficient of variation for the unit *i* and the sign ± depends on the kind of phenomenon to be measured (De Muro *et al.*, 2010).

To facilitate the interpretation of results, we suggest to choose the 'goalposts' so that 100 represents a reference value (e.g., the average in a given year).

A simple procedure for setting the 'goalposts' is the following.

Let Inf_{x_j} and Sup_{x_j} be the overall minimum and maximum of the indicator *j* across all units and all time periods considered. Denoting with Ref_{x_j} the reference value for the indicator *j*, the 'goalposts' are defined as:

$$\begin{cases} \text{Min}_{x_j} = \text{Rif}_{x_j} - \Delta \\ \text{Max}_{x_j} = \text{Rif}_{x_j} + \Delta \end{cases}$$

where $\Delta = (\text{Sup}_{x_j} - \text{Inf}_{x_j})/2$. The normalized values will fall approximately in the range (70; 130), where 100 represents the reference value.

The AMPI allows to compare the trends of the various units with the average trend, in contrast to the MPI where the average value is set to 100 for each time period. In addition, it may be simultaneously applied to different type of units (e.g., countries, regions, cities) without loss of comparability.

The 'price' to pay for having scores comparable over time is that individual indicators with different variability are aggregated. However, normalized indicators in an identical range have much more similar variability than original ones.

[3] The 'polarity' of a individual indicator is the sign of the relation between the indicator and the phenomenon to be measured (+ if the individual indicator represents a dimension considered positive and – if it represents a dimension considered negative).

3. An application to socio-environmental data

In order to test the method presented in the previous Section, a set of socio-economic and environmental indicators were selected according to the capability approach theory (Sen, 1985). In fact, the pillars considered are 'health', 'wealth distribution', 'work', 'education' and 'environment'. We based on the theory underlying the HDI, to which we added the pillars 'work' and 'environment'. This choice is motivated by the fact that the development index must interpret the phenomenon for the Italian regions e not for developing countries.

The individual indicators selected are: I_1) 'Life expectancy at birth', expressed in years (positive polarity); I_2) 'Income distribution inequality' - Gini coefficient (negative polarity); I_3) 'Employment rate for people aged 20-64', expressed in percentage (positive polarity); I_4) 'People aged 25-64 with low education level', expressed in percentage (negative polarity); I_5) 'Greenhouse gas emissions', expressed in CO_2 equivalent tons per capita (negative polarity). Since the phenomenon to be measured is 'positive', it is necessary to normalize the individual indicators that have negative polarity by calculating the complement of (1) with respect to 200.

In order to make a comparison over time, two years were selected: 2004 and 2011. In Table 1 are reported the original values of the five individual indicators.

Table 1 – *Individual indicators of development in the Italian regions - Years 2004, 2011.*

Region	2004					2011				
	I_1	I_2	I_3	I_4	I_5	I_1	I_2	I_3	I_4	I_5
Piemonte	80.6	0.309	66.9	52.0	9.8	81.8	0.303	68.4	42.7	7.1
Valle d'Aosta	80.6	0.296	70.7	54.9	6.8	81.8	0.282	71.2	48.3	4.9
Liguria	80.9	0.314	63.5	44.2	12.3	81.6	0.341	67.4	37.1	9.1
Lombardia	81.0	0.320	69.1	49.3	9.6	82.3	0.291	69.0	41.6	8.4
Bolzano/Bozen	81.2	0.298	73.0	58.1	6.1	83.2	0.256	76.0	46.3	5.5
Trento	81.2	0.271	69.6	43.3	6.1	82.8	0.274	71.0	34.2	5.5
Veneto	81.3	0.281	67.7	53.6	10.2	82.4	0.276	69.2	42.8	7.7
Friuli-V.G.	80.6	0.273	65.8	49.0	11.6	81.7	0.301	68.2	42.1	10.6
Emilia-Romagna	81.3	0.299	71.7	48.0	12.2	82.4	0.289	72.1	39.4	9.9
Toscana	81.6	0.268	66.8	51.7	7.6	82.6	0.283	67.6	45.0	5.9
Umbria	81.5	0.286	65.2	43.3	14.0	82.6	0.278	66.6	34.1	9.9
Marche	81.9	0.280	67.8	48.5	7.0	82.9	0.284	67.2	42.1	6.4
Lazio	80.2	0.328	62.6	41.6	7.7	81.8	0.328	63.2	33.9	6.4
Abruzzo	81.0	0.293	60.7	47.0	5.8	82.1	0.279	61.1	38.4	4.1
Molise	81.0	0.286	56.4	51.2	8.3	82.1	0.303	54.7	47.5	7.8
Campania	79.4	0.347	49.2	57.7	3.6	80.4	0.353	43.1	52.9	3.7
Puglia	81.2	0.303	48.8	60.4	14.1	82.1	0.314	48.6	54.1	11.9
Basilicata	80.5	0.298	53.6	53.0	4.7	82.0	0.344	51.7	46.1	2.9
Calabria	80.8	0.333	50.5	53.5	3.4	82.1	0.317	46.2	48.4	3.2
Sicilia	80.2	0.348	47.0	59.5	8.4	81.1	0.334	46.2	53.2	7.7
Sardegna	80.8	0.323	55.0	61.4	11.6	81.9	0.277	55.6	53.5	9.5
Italia	**80.8**	**0.328**	**61.3**	**51.9**	**8.9**	**82.0**	**0.319**	**61.2**	**44.3**	**7.4**

The choice of the years is due to both the availability of data on a large series and the interest of the comparison between a time before the crisis and a time of crisis. As explained in Section 2, the AMPI allows to make comparisons over time without any methodological problem. The 'goalposts' were calculated so that 100 represents the Italy's value in 2004.

The results are shown in Table 2, where for each of the two years and for each Italian region the AMPI is reported together with its rank. In the last two columns, the differences of scores and ranks are computed. The first thing to note is that, consistent with the assumptions of the model, all values fall in the range (70; 130) and the value 100 is set for Italy in 2004. Considering temporal differences, the socio-economic-environmental development in Italy increased, between 2004 and 2011, more than 9 points as well as all regions showing differences in a positive sign: from +1.75 of Campania to +18.57 of Bolzano. So, the important message is that, compared to 2004, Italy and its regions seem to be more developed from the social, economic and environmental point of view.

Table 2 – *Composite Index of development in the Italian regions - Years 2004, 2011 and variation 2004-2011.*

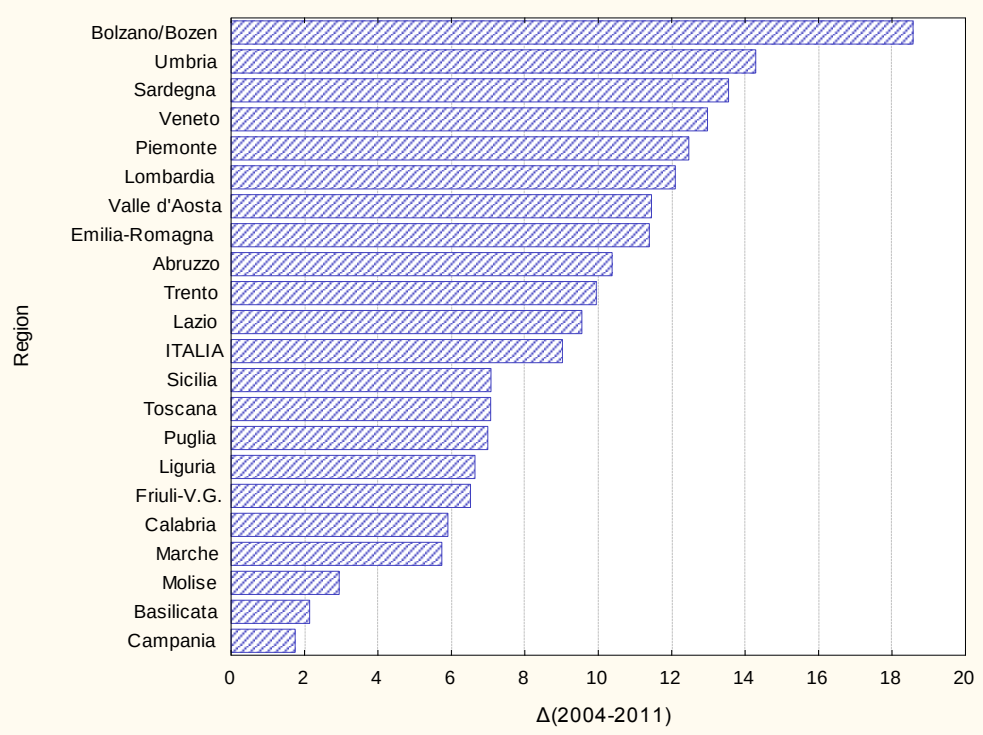
Region	2004		2011		Δ(2004-2011)	
	Value	Rank	Value	Rank	Value	Rank
Piemonte	102.06	14	114.53	11	12.47	3
Valle d'Aosta	106.33	6	117.79	8	11.45	-2
Liguria	100.91	16	107.56	14	6.65	2
Lombardia	104.43	9	116.53	10	12.10	-1
Bolzano/Bozen	108.04	5	126.61	2	18.57	3
Trento	117.13	1	127.07	1	9.95	0
Veneto	105.90	7	118.88	6	12.98	1
Friuli-V.G.	103.54	12	110.06	13	6.52	-1
Emilia-Romagna	105.35	8	116.74	9	11.39	-1
Toscana	112.08	3	119.16	5	7.07	-2
Umbria	103.69	10	117.99	7	14.29	3
Marche	114.62	2	120.36	3	5.74	-1
Lazio	103.04	13	112.60	12	9.55	1
Abruzzo	109.37	4	119.74	4	10.38	0
Molise	103.57	11	106.52	15	2.95	-4
Campania	87.82	19	89.57	21	1.75	-2
Puglia	87.27	20	94.26	19	6.99	1
Basilicata	101.75	15	103.89	16	2.14	-1
Calabria	97.81	17	103.72	18	5.91	-1
Sicilia	86.50	21	93.58	20	7.08	1
Sardegna	90.34	18	103.89	17	13.55	1
Italia	**100.00**		**109.03**		**9.03**	

However, the intensity of development does not seem to be equal and constant for the Italian regions because, from the differences of rank, we note that there are negative signs. This means that, even if all regions improve over time, the 'speed' of improvement is different and several crossovers in the ranking are possible. For

example, the region Lombardy, although increased the value of the composite index of more than 12 points, dropped one place in the ranking.

In Figure 1, the composite indices' variations, between 2004 and 2011, are presented. The composite index of Bolzano increased by more than 18 points and has the best acceleration, followed by Umbria with over 14 points. Note that Sicilia, Toscana, Puglia, Liguria, Friuli, Calabria, Marche, Molise, Basilicata and Campania increased less than the Italian average.

Figure 1 – *Composite Index of development in the Italian regions - Variation 2004-2011.*



The application to real data demonstrated the goodness of the proposed method that enables to make comparisons over time and space. It is good to consider, also, that the composite index cannot 'invent' an information that is not contained in the individual indicators of the original matrix. The composite index must well understand the 'message' of multidimensional reality, trying to lose less information as possible.

4. Conclusions

The design and the implementation of a composite index is a complex process that involves well-defined steps of work, where the arbitrary choices of the researcher have a significant effect on the final results. The heated debate within the scientific Community, over the years, seems to converge towards the idea that there is not a composite index universally valid for all areas of application, and, therefore, its validity depends on the strategic objectives of the research (Mazziotta and Pareto, 2013a).

Usually, when the phenomenon to be measured is the well-being, it is necessary, for a correct use and communicability of the results, that the composite index satisfy the following requirements: a) spatial comparability, i.e., the possibility to compare composite values between territorial units; b) comparability over time, i.e., the possibility to compare composite values over time; c) 'non-substitutability' of the individual indicators, i.e., the attribution of the same weight to the individual indicators and the inability to compensate the value of one with that of another; d) simplicity and transparency of the calculation; e) immediate use and interpretation of the output results; f) robustness of the results.

The AMPI meets all the requirements because the methods of normalization and aggregation are chosen so that the research objectives are achieved. Requirements a) and b) are met since the *min-max* normalization produces absolute values in the range (70; 130) where the reference value is set at a given time *t*; requirement c) is met since the penalty function does not allow a full compensation of the indicators; requirement d) is met since the method is based on the arithmetic mean and the standard deviation so that anyone can easily reproduce the results; requirements e) is met since the reference value is fixed equal to 100 at a given time *t* and so it is very easy to identify immediately the units that are above and below that value; requirement f) is met since the 'sensitivity analysis' carried out among several alternative methods shows that the MPI is always the most robust composite index (Mazziotta C. *et al.*, 2010).

References

BANDURA R. 2008. *A Survey of Composite Indices Measuring Country Performance: 2008 Update*. New York: UNDP/ODS Working Papers.

GNESI C., SEGRE E., VILLA A. 2010. *Come si vive in Italia? Rapporto QUARS 2010 – Indice di qualità regionale dello sviluppo*. Roma: Sbilanciamoci!

ISTAT 2014. *Bes 2014. Il benessere equo e sostenibile in Italia*. Roma: Istat.

LUN G., HOLZER D., TAPPEINER G., TAPPEINER U. 2006. The stability of rankings derived from composite indicators: analysis of the "Il Sole 24 ore" quality of life report. *Social Indicators Research*, Vol. 77, No. 2, pp. 307-331.

DE MURO P., MAZZIOTTA M., PARETO A. 2010. Composite Indices of Development and Poverty: An Application to MDGs. *Social Indicators Research*, Vol. 104, No. 1, pp. 1-18.

MAZZIOTTA C., MAZZIOTTA M., PARETO A., VIDOLI F. 2010. La sintesi di indicatori territoriali di dotazione infrastrutturale: metodi di costruzione e procedure di ponderazione a confronto. *Rivista di Economia e Statistica del Territorio*, Vol. 1, pp. 7-33.

MAZZIOTTA M, PARETO A. 2013. Methods for constructing composite indices: one for all or all for one? *Rivista Italiana di Economia Demografia e Statistica*, Vol. LXVII, No. 2, pp. 67-80.

MAZZIOTTA M., PARETO A. 2013. A Non-compensatory Composite Index for Measuring Well-being over Time. *Cogito. Multidisciplinary Research Journal*, Vol. V, No. 4, pp. 93-104.

OECD 2008. *Handbook on Constructing Composite Indicators. Methodology and user guide*. Paris: OECD Publications.

SEN A. K. 1985. *Commodities and Capabilities*. Amsterdam: Elsevier.

TARANTOLA S. 2008. *European Innovation Scoreboard: strategies to measure country progress over time*. Luxembourg: Publications of the EC.

UNDP 2010. *Human Development Report 2010. The Real Wealth of Nations: Pathways to Human Development*. New York: Palgrave Macmillan.

SUMMARY

Most of the socio-economic phenomena such as development, well-being or societal progress have a multidimensional nature and require the definition of a set of individual indicators in order to be properly assessed. Often, individual indicators are summarized and a composite index is created. One of the main problems in constructing composite indices is the choice of a method that allows to assess changes over time. In this paper, we present a variant of the Mazziotta-Pareto Index, namely Adjusted MPI, which allows time comparisons across units to be made. An application to a set of indicators of socio-environmental development in the Italian regions is presented.

Matteo MAZZIOTTA, Italian National Institute of Statistics, mazziott@istat.it
Adriano PARETO, Italian National Institute of Statistics, pareto@istat.it

Rivista Italiana di Economia Demografia e Statistica Volume LXVIII n. 3/4 Luglio-Dicembre 2014

A NONPARAMETRIC GINI CONCENTRATION TEST FOR LABOUR MARKET ANALYSIS

Chiara Gigliarano, Francesco Maria Chelli

1. Introduction

The Gini index is one of the most important statistical indices employed in social sciences for measuring concentration in the distribution of a positive random variable; it is mainly used in economics as a measure of income or wealth inequality among individuals or households (see, e.g., Gini 1912, 1914). Recently, the Gini coefficient has been used to describe concentration in levels of mortality, or in length of life, among different socio-economic groups, and to evaluate inequality in health and in life expectancy (see, e.g., Hanada 1983; Bonetti et al. 2009).

Aim of this paper is to analyse the differences in survival times of job contracts among subgroups of workers, from the point of view of concentration.

We examine the differences both in the length of the first job contract and in the waiting time between the end of the first contract and the beginning of a new one.

We apply the well known Gini index to measure concentration in survival times within groups of workers, and as a way to compare the distribution of survival times across such groups. We consider a test for differences in the heterogeneity of survival distributions, which may suggest the presence of a differential covariates effect on the job contract survival.

The analysis is based on the Italian Compulsory Communications system data, which record all the activations, transformations, fixed-term extensions and anticipated terminations of employment relationships between any worker and employer in Italy since January 2009 until June 2012. The target population is made up by the young workers, between 18 to 35 years old.

The rest of the paper is structured as follows: in Section 2 we briefly review the Gini test for survival data; in Section 3 we analyse the Italian labour market from the point of view of concentration; in Section 4 we conclude.

2. The Gini index for survival data: a brief review

The Gini index measures concentration in the distribution of a positive random variable. Bonetti et al. (2009) propose to apply the Gini index in survival analysis in order to measure concentration in survival times within groups of subjects. In particular, they apply a restricted version of the Gini index to right-censored survival data in order to detect differences in concentration (heterogeneity) between the survival time distributions of two groups.

A number of nonparametric statistical tests exist in the literature to test the difference in survival distribution functions between groups. Common tests are in the class of weighted linear rank tests, including the log-rank test (LR test), the Wilcoxon test (W test), the Gray and Tsiatis test (GT test); see, e.g., Harrington and Fleming 1982; Gray and Tsiatis 1989. Testing for differences between survival distributions via a concentration measure may prove more powerful than these methods, for example when one is far from the proportional hazard structure.

The Gini coefficient of concentration for a positive random variable X with cumulative distribution function *F* and survival function *S* is defined as

$$G = 1 - \frac{\int_0^\infty [1 - F(x)]^2 dx}{\int_0^\infty \Pr(X > x)\, dx} = 1 - \frac{\int_0^\infty [S(x)]^2 dx}{\int_0^\infty S(x) dx} \; ;$$

see Hanada, 1983.

In survival analysis subjects have usually a finite follow-up time, so we consider the restricted version of the Gini index:

$$G_t = 1 - \frac{\int_0^t [S(x)]^2 dx}{\int_0^t S(x) dx} \; ,$$

where *t* represents the longest follow-up time in the data.

Minimum value of G_t is reached when all subjects have the same survival time, while maximum value is obtained when one individual has the maximum survival time and the rest of the population experiences the event immediately.

Bonetti et al. (2009) and Gigliarano and Bonetti (2013) propose a test based on the restricted Gini index G_t for comparing two survival functions related to two different groups. Their Gini test is aimed to test for differences in two survival distributions from the point of view of concentration. The Gini test statistic is

$$T = \frac{\left(\hat{G}_{1,t} - \hat{G}_{2,t}\right)^2}{\widehat{Var}(\hat{G}_{1,t}) + \widehat{Var}(\hat{G}_{2,t})}$$

where $\hat{G}_{j,t}$ is the estimator of the restricted Gini index for censored data referred to the group *j* and $\widehat{Var}(\hat{G}_{j,t})$ is the estimator of the approximate variance of $\hat{G}_{j,t}$, for group *j* , *j* = 1, 2.

Bonetti et al. (2009) prove that under the null hypothesis of equality of the two survival distributions, the statistic *T* has an approximate chi-squared distribution with 1 degree of freedom, while, under any alternative to the null hypothesis, *T* is distributed as an approximate noncentral chi-squared distribution.

3. Data description

The empirical illustration is based on a sample of the Compulsory Communications ("Comunicazioni Obbligatorie") data provided by Italian Ministry of Labour and Social Policies.[1]

The Compulsory Communications (henceforth, CC) data include all activations, transformations, fixed-term extensions, early anticipated terminations of a working relationship, either public or private.

The sample refers to all Italian workers born on 15 January, 15 April, 15 July and 15 October of any year. Our database therefore includes about 1 out of 91 of all workers who have been involved in the CC system over the period between January 2009 and June 2012.

The population of interest are the 18-35 aged workers who activated a contract in 2009. Individuals who entered the CC database for the first time after December 31, 2009 are excluded from the analysis.

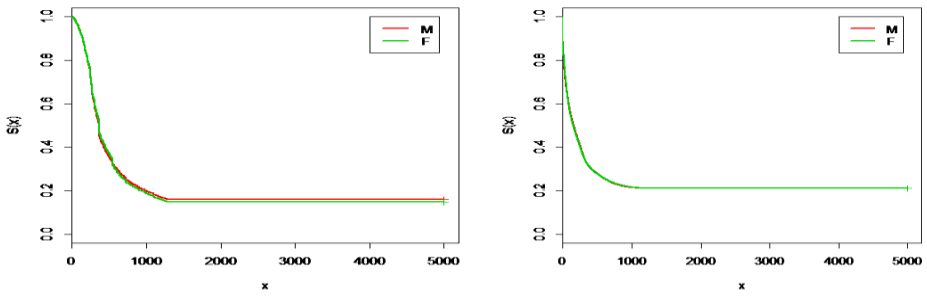
The CC data have as unit of observation the contract ("contratto di lavoro"), defined as a working relationship between an employer and an employee and characterized by a starting date. However, in the context of mobility analysis, the key concept is the worker rather than the contract; therefore, the worker's history needs to be reconstructed starting from the original CC data, so that the observation unit becomes the individual.

For more details on the data preparation and cleaning process we refer to Lilla and Staffolani (2011), while further information on the methodology for joining

[1] The Compulsory Communication Data are used with the permission of the Ministry of Labour and Social Policies thanks to the agreement between the Department of Economics and Social Sciences of Marche Polytechnic University and General Department for the Innovation Technology of the Ministry of Labour and Social Policies. The authors are grateful to Stefano Staffolani and Matteo Picchio for the data preparation.

different contracts corresponding to same individual can be found in Picchio and Staffolani (2013).
CC data provides information on the daily occupational status of an individual. Here for simplicity a monthly unit of time is considered, and for each month he *prevalent* contract is selected (according to type and length of contract).
The variable of interest is the occupational status. Four are the types of occupational status considered, that are ordered as follows: (i) *not in employment*, (ii) *temporary contract*, including fixed-term contract ("contratto a tempo determinato"), parasubordinate contract ("contratto di collaborazione coordinata e continuativa"), internship contract ("contratto di stage"), interim contract ("lavoro interinale"), (iii) *apprenticeship contract* ("contratto di apprendistato"), (iv) *permanent contract*, that is the open-ended contract ("contratto a tempo indeterminato").
We apply the Gini test discussed above to the measurement of concentration in survival times within groups of workers, and as a way to compare the distribution of survival times across such groups.
Analysis of the differences in survival times of job contracts has been performed among subgroups of workers, based on gender, educational level and geographical area.
In particular, we have analysed differences both *(i)* in the length of the first job contract and *(ii)* in the waiting time between the end of the first contract and the beginning of the second one. The results are summarised in Table 1 and illustrated in Figures 1 to 4.

Figure 1 – *Male versus female. Left-hand side: Length of the first job. Right-hand side: Waiting time for a new first job.*



A first analysis is aimed at determining whether there are gender differences in the Italian labour market. Figure 1 and Table 1 reveals that there exists no significant difference between young males and young females in the waiting time between

the end of the first contract and the beginning of a new one, while significant differences emerge in the length of the first job contract, which is longer for males and females.

Table 1 – *P-values of Gini, Gray-Tsiatis (GT), Log Rank (LR) and Wilcoxon (W) tests for different groups comparisons.*

		Gini	GT	LR	W
GENDER (Male versus female)	Length of the first job	0.0152	0.0051	0.4041	0.4997
	Waiting time for new job	0.8366	0.7629	0.9687	0.9865
EDUCATION (Tertiary versus non tertiary)	Length of the first job	0.0000	0.4646	0.0000	0.0000
	Waiting time for new job	0.0000	0.0000	0.0000	0.0000
EDUCATION IN TERTIARY SECTOR (Tertiary versus non tertiary)	Length of the first job	0.0000	0.0000	0.5114	0.0463
	Waiting time for new job	0.0000	0.0000	0.0000	0.0000
GEOGRAPHICAL AREA (North versus South)	Length of the first job	0.8834	0.0000	0.0000	0.0000
	Waiting time for new job	0.0000	0.0000	0.0000	0.0000

We also test for the presence of significant impact of the educational level on the Italian labour market: Table 1 and Figure 2 shows that tertiary education helps in finding quickly a new job, while it seems not so relevant for activating permanent contracts. With a particular focus on the tertiary economic sector, if a worker has tertiary education he will find quicker a job at the end of the first contract, but the length of his first contract will be shorter, in comparison to workers in the same economic sector but without tertiary education (see Table 1 and Figure 3).

Figure 2 – *Tertiary education versus non-tertiary education. Left-hand side: Length of the first job. Right-hand side: Waiting time for a new job.*

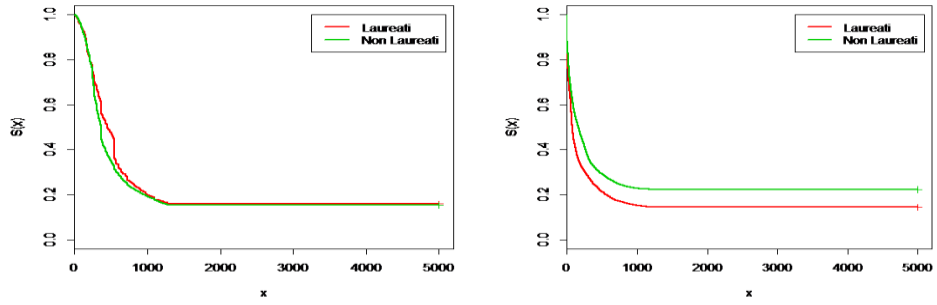
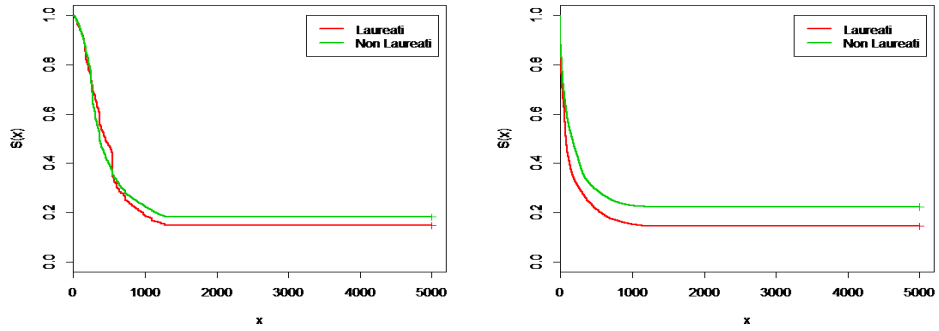
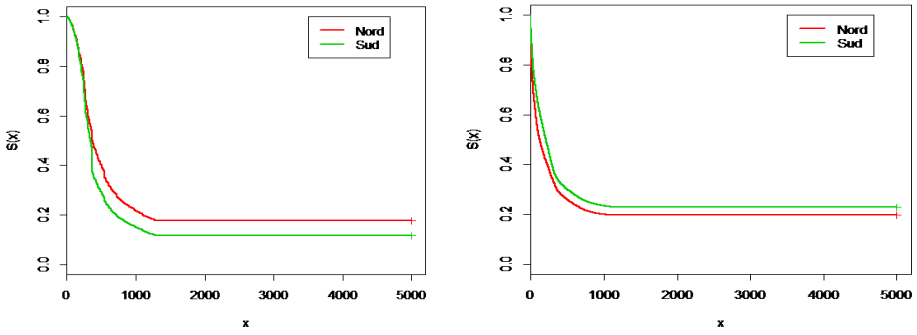


Figure 3 – *Tertiary education versus non-tertiary education within the tertiary economic sector. Left-hand side: Length of the first job. Right-hand side: Waiting time for a new job.*



Finally, we compare the Italian macro areas (North, Center and South): no statistically significant differences emerge between North and Center of Italy (data are not shown), while differences emerge between North (or Center) and South of Italy. Table 1 and Figure 4 reveals that the labour market in the North of Italy is characterized by higher percentage of permanent contracts and by shorter waiting time for the activation of the second contract, if compared to the South of Italy.

Figure 4 – *North versus South of Italy. Left-hand side: Length of the first job. Right-hand side: Waiting time for a new job.*



4. Concluding remarks

In this paper we have examined the Italian labour market dynamics from a novel point of view, based on the concentration analysis.

The empirical analysis revealed that there exists no significant difference between male and female in the waiting time between the end of the first contract and the beginning of a new one. Gender differences emerge, instead, in the length of the first job contract, which appears to be significantly longer for males than for females.

Significant differences emerge also among geographical areas: the North of Italy has the highest percentage of permanent contracts and also the shortest waiting time for the second contract.

Finally, different levels of education have different impact on the Italian labour market: tertiary education helps in finding quickly a new job, while it seems not so relevant for activating permanent contracts.

References

BONETTI M., GIGLIARANO C., MULIERE P.,2009. The Gini concentration test for survival data. *Lifetime Data Analysis*, Vol. 15, pp. 493-518.

GIGLIARANO C., BONETTI M. (2013), Gini test for survival data in presence of small and unbalanced groups, Epidemiology, Biostatistics and Public Health, Volume 10, Number 2, DOI:10-2427/8762.

GINI C. (1912) Variabilità e mutabilità. Contributo allo studio delle distribuzioni e relazioni statistiche. Studi Economico-Giuridici dell'Università di Cagliari III

GINI C. (1914) Sulla misura della concentrazione e della Variabilità dei caratteri. Atti del Reale Istituto Veneto di Scienze, Lettere ed Arti LXXIII(part 2):1203–1248.

GRAY R.J., TSIATIS A.A., 1989. A linear rank test for use when the main interest is in differences in cure rates, *Biometrics*, Vol. 45, pp.899-904.

HANADA K., 1983. A formula of Gini's concentration ratio and its applications to life tables, *Journal of the Japan Statistical Society*, Vol. 19, pp.293-325.

HARRINGTON D.P., FLEMING T.R., 1982. A class of rank test procedures for censored survival data, *Biometrika* Vol. 69, No. 3, pp.553-566.

PICCHIO M. and STAFFOLANI S. 2013. Does Apprenticeship Improve Job Opportunities? A Regression Discontinuity Approach, *IZA DP No. 7719.*

SUMMARY

Aim of the paper is to analyse the differences in survival times of job contracts among subgroups of workers, based on age, gender, educational level, region.

We examine the differences both in the length of the first job contract and in the waiting time between the end of the first contract and the beginning of a new one.

We apply the well known Gini index to the measurement of concentration in survival times within groups of workers, and as a way to compare the distribution of survival times across such groups. We consider a test for differences in the heterogeneity of survival distributions, which may suggest the presence of a differential covariates effect on the job contract survival.

The analysis is based on the Italian Compulsory Communications system data for the period between January 2009 and June 2012.

Chiara GIGLIARANO, Dipartimento di Scienze Economiche e Sociali, Università Politecnica delle Marche, Ancona, e-mail: c.gigliarano@univpm.it

Francesco Maria CHELLI, Dipartimento di Scienze Economiche e Sociali, Università Politecnica delle Marche, Ancona, e-mail: f.chelli@univpm.it

FERTILITY DYNAMICS IN EUROPE: REFLECTIONS ON THE PRINCIPAL INTERPRETATIVE PARADIGMS IN LIGHT OF SOME EMPIRICAL EVIDENCE[1]

Barbara Zagaglia, Eros Moretti

1. Introduction

Since the second half of the last century, European countries have undergone deep demographic changes, continuing along a path that they started more than a century previously. These transformations mainly concern reproductive behaviours, although major improvements have also been made in survivorship and important changes have affected migration flows.

The transformations have been so important that some scholars consider them to mark a new and autonomous path and interpretative scheme. We refer to the Second Demographic Transition (SDT) theory, which since it appeared in the second half of the 1980s, has been criticized on various grounds and to different extents (for a review, see, for instance, Lesthaeghe "Second Demographic Transition", Basil Blackwell, and Lesthaeghe, 2010). At the same time, numerous explanations, not in opposition to SDT theory, have been put forward to account for a fertility decline below the replacement level and its further fall to lowest-low levels (for a review see, for instance, Zagaglia, 2006).

The SDT theory has developed in recent years, and it has been enriched with new investigations. Numerous doubts in its regard have been dispelled, and, today it is the prevalent explanatory paradigm in Europe.

In this paper, we consider the long-term evolution of the main dimensions of fertility in order to reflect on the theoretical bases of the dominant theory as it has recently developed. Owing to the limited amount of space available, we restrict the analysis and the consequent discussion to what we consider to be the main aspects of the theory, and we select countries according to the representativeness criterion.

[1] Sections 1 and 4 are to be attributed to both authors, sections 2 and 3 to Barbara Zagaglia.

2. The present demographic regime according to the SDT paradigm

According to the SDT theory, the main characteristics of the present demographic regime are sustained sub-replacement fertility and a multitude of living arrangements other than marriage. Procreation is disconnected from marriage and relies on perfect contraception which is widely and socially accepted. (Lesthaeghe, "Second Demographic Transition", Basil Blackwell).

To focusing on fertility - the most salient and least debated aspect of the SDT theory - the explanation for fertility decline below replacement level is the postponement of parenthood and the shifting of fertility to older ages (Lesthaeghe, 2010), while the ultimate and common cause of both of them are new life-style preferences, in particular value orientation towards individual autonomy and self-realization. Taken as a whole, the present demographic regime can be summarized as a low(er) and late(r) fertility regime[2] (Lesthaeghe, 2006).

New values have been recognized as responsible for the shift to the new regime since the original formulation of the SDT theory by Lesthaeghe and van de Kaa (1986) and van de Kaa (1987). Recently, van de Kaa (2002) and Sobotka (2008) have demonstrated that those new values are good predictors of postponement.

We reproduce below the figure from Sobotka (Figure 1) which demonstrates the negative relationship between an index of typical SDT values and the date of onset of the postponement of the first birth.

The different timing of postponement reflects the European diffusion of the SDT from Northern and Western European countries to Southern and Central and Eastern Europe. In this regard, a first formulation of the theory (van de Kaa, 1987), which posited a sequence of four distinct phases with which the other European countries after Northern and Western countries could conform and through which they could pass[3], was later replaced by new evidence. In the 1990s, features of the SDT arose in the rest of Europe and, outside Europe, in Canada, Australia, New Zealand, the USA, Japan, South Korea, Taiwan, Hong Kong and Singapore, giving rise to multiple patterns or variants (Lesthaeghe and Neider, 2006 and Lesthaeghe, 2010).

According to Lesthaeghe (2010), in Europe, after the collapse of the Communist regime all the SDT features emerged simultaneously in Central and Eastern countries. Extra-marital fertility had already started to increase before the collapse (Lesthaeghe, 2010), but it greatly expanded after 1989, together with the postponement of childbearing and a rapid decline in fertility to very low rates (Sobotka, 2004).

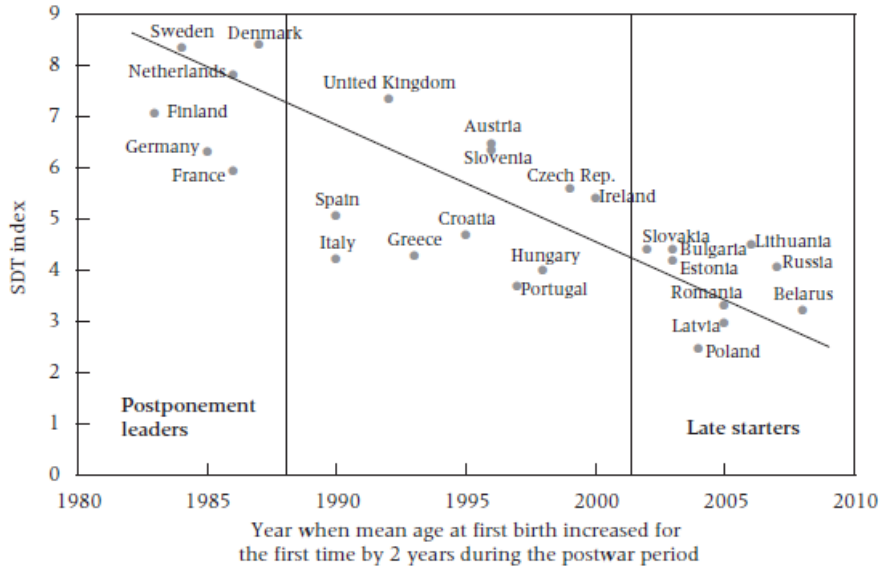
[2] Lower than the level of substitution and later than in the transitional regime.
[3] This aspect was much disputed in the 1990s.

In Southern Europe, some SDT characteristics, such as the spread of alternative family forms and extra-marital births, were low for a long time, and residence in the parental home was prolonged. Recently, however, both cohabitation and births in informal unions have spread in many of these countries as well. Fertility and marriage postponement started late, but they were more noticeable and stronger than in Northern and Western Europe.

In these regions the recovery of delayed births at later ages has been weak, so that a pattern of lowest-low fertility has emerged in them.

Finally, as regards the countries which first entered the SDT and experienced all its features to the greatest extent, the SDT has developed further. In these countries, the proportions of births out of wedlock have further increased, and in Iceland, Sweden, Germany, Norway, and France, they are now extremely high. Northern and Western women postponed motherhood first and their fertility declined early, but a considerable recovery at older ages has now enabled them to reach high levels of sub-replacement fertility (Lesthaeghe, 2010).

Figure 1 – *Relationship between a composite index of SDT values in 1999-2000 and the date of the onset of the postponement of the first birth*



Source: Sobotka (2008). The SDT index is from Sobotka (2008, pp.86-87). In Lesthaeghe (2010).

3. Long-term analysis of the timing-intensity relationship in some countries representative of the different variants of SDT

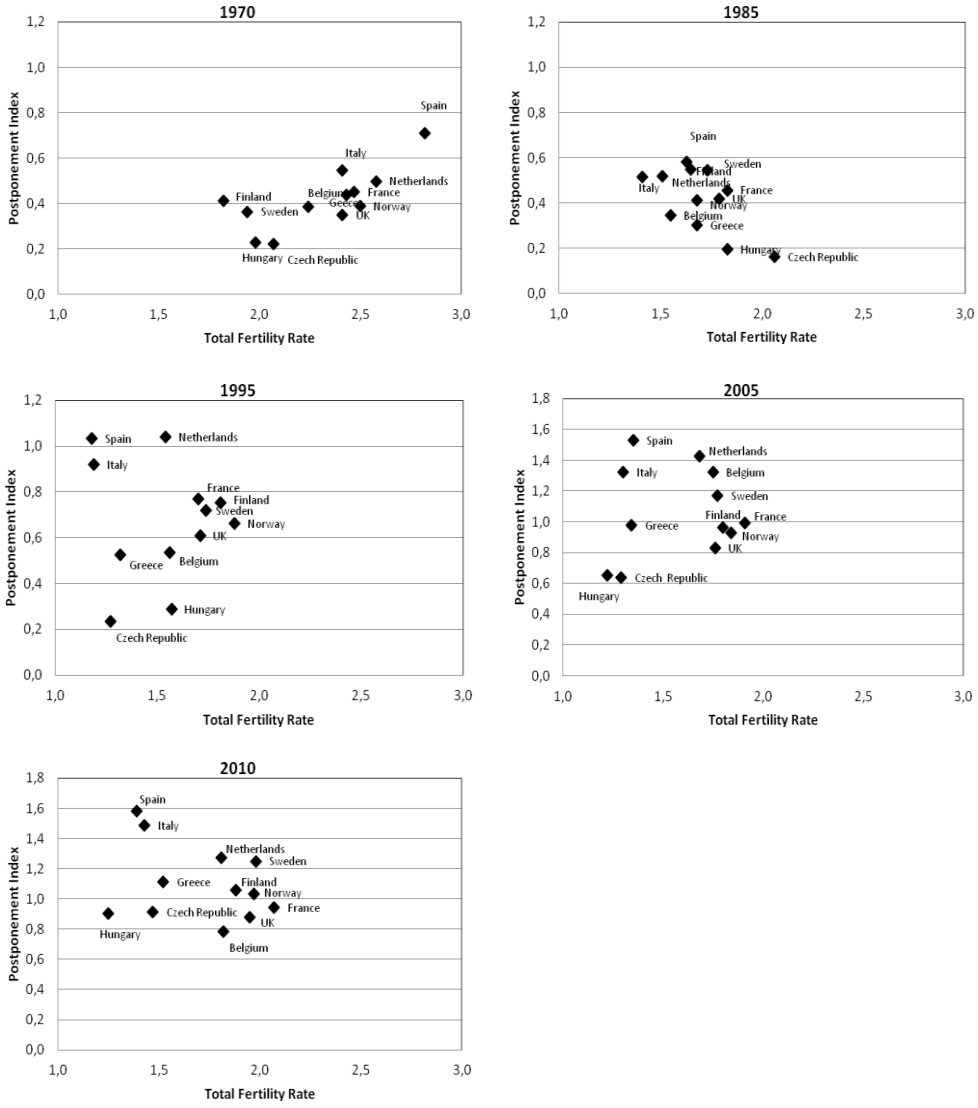
In this section we investigate the relationship between the shift of motherhood to older ages and fertility decline in a long-term perspective.

Indeed, whereas the relationship between SDT values and fertility postponement and the relationship between SDT values and fertility levels have been tested (Sobotka, 2008), a real test of the relationship between fertility postponement and fertility level is, to the best of our knowledge, still lacking in the literature. Moreover, analysis has usually concerned limited periods of time.

First, in line with the literature (Sutton and Mathews, 2004), we consider, as an index of postponement as well as a measure of cadence, the ratio between live births to women aged over thirty and those to younger women (less than thirty years old). The correlation between this index and the Total Fertility Rate (TFR) is shown in Figure 2 for selected countries and years.

In 1970, the correlation between the intensity and the cadence of fertility was positive (ρ=0.7203) and produced by two distinct scatters (split correlation). In the case of over-replacement fertility, higher intensities were associated with higher levels of the postponement index, while in the case of sub-replacement fertility, lower intensities were associated with higher values of the index of postponement. A negative correlation between the fertility decline and motherhood postponement was clearly apparent in the mid-1980s (ρ=-0.6213), when all the countries selected had fertility levels below replacement. According to our analysis, the negative and strong correlation has weakened since the mid-1990s (ρ equals to -0.0682, 0.1091, and -0.3120 in 1995, 2005 and 2010, respectively), as the SDT has spread to Central, Eastern, and Southern Europe, and it has happened earlier than stated by Lesthaeghe (2010). Here, due to the limited space available, we focus only on some countries selected in order better to represent the different variants of the SDT. However, similar results hold when the full set of European countries is considered.

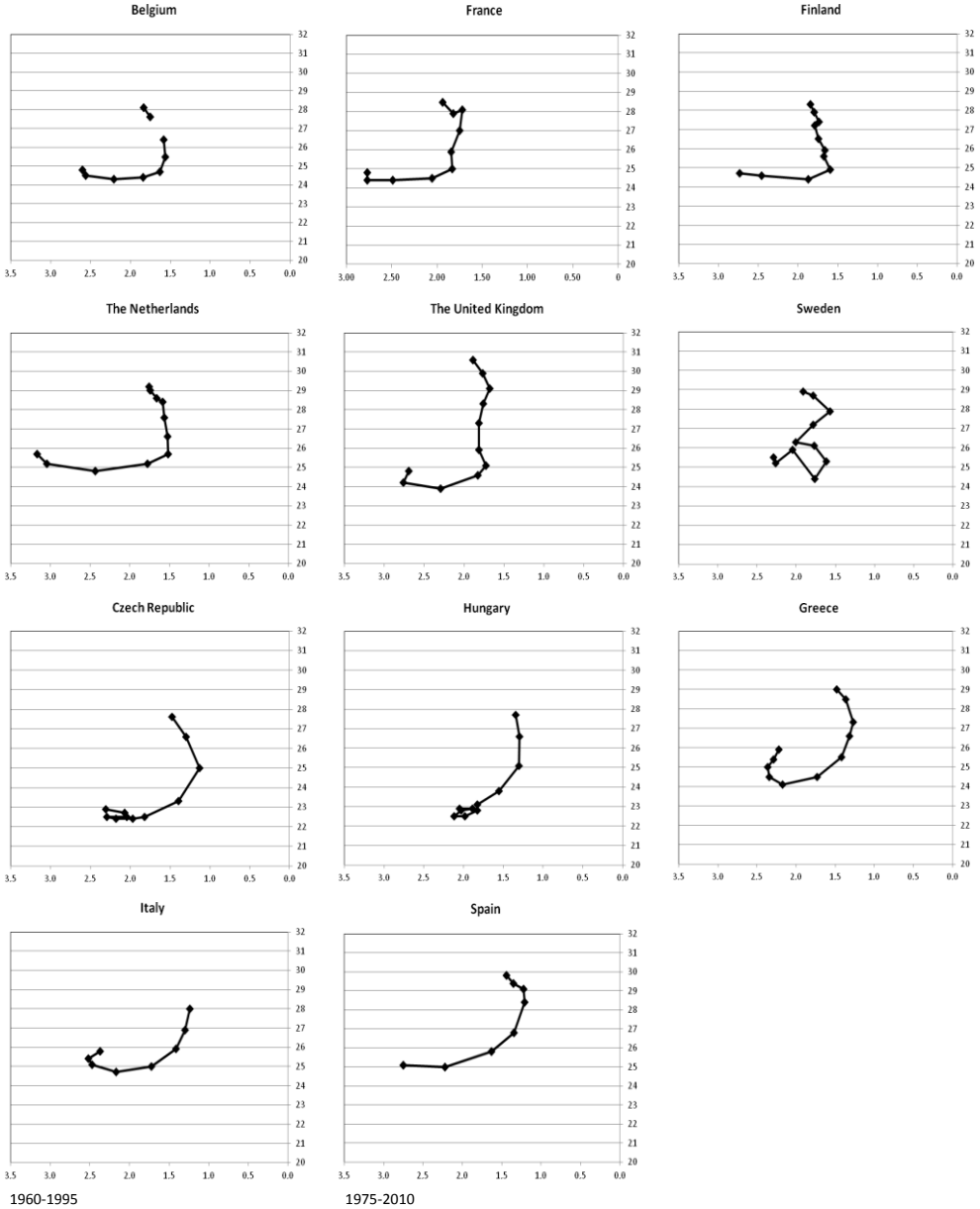
Figure 2 – *Relationship between TFR and the Postponement Index. 1970-2010.*



Source: Our elaborations on World Fertility Data 2012, United Nations, 2013.

Moreover, because at high intensities of fertility the postponement index may not show a change in the timing of childbearing owing to the high proportion of high-order births to women aged over 30, we consider a different and more robust indicator of fertility postponement: the mean age at first birth (MAFB).

Figure 3 – *Relationship between TFR and Mean age at first birth. 1960-2010.*



Source: Our elaborations on United Nations, World Population Prospects: The 2012 Revision, 2013, and Eurostat database.

Figure 3 shows the joint path of MAFB and TFR for individual countries[4]. Two different patterns are apparent. One pattern, for the first SDT movers (Northern and Western countries), is characterized by a dissociation between the fertility decline and the delay of motherhood which contrasts with what the SDT theory states. In Belgium, France, Finland, the Netherlands, the United Kingdom, fertility fell below replacement level without postponement, which, instead, to a larger extent appeared later, when fertility was below replacement. A second pattern, by contrast, is characterized by changes (decreases) in the fertility intensity associated with changes (increases) in the timing of fertility, in accordance with the SDT theory. This pattern is exhibited by Czech Republic, Hungary, Greece, Italy and Spain, Central, Eastern, and Southern European countries and late comers in the SDT.

5. Final remarks

We have focused on the demographic aspect of the prevalent paradigm explaining the persistent low fertility in Europe, and we have analyzed the long-term relationship between the level and the timing of fertility. The results challenge the causal relationship between the postponement of childbearing and fertility decline that is a central feature of the SDT theory. In this regard, the Northern and Western pattern proves not to be representative for the SDT. According to our findings, the salient SDT characteristics on fertility fit better with the *late comers* than the *first movers.* Moreover, the recent and long-lasting dichotomy between lowest-low and highest-low fertility (where the distinction is a TFR lower or higher than 1.5), which is unrelated to changes in the cadence of fertility, weakens the explanatory power of the SDT paradigm in favor of other theories. In particular, institutional perspectives seem better to explain the most recent dynamics of fertility. Indeed, the role played by the social and economic structures, together with the policies implemented, may be crucial in orienting the fertility choices of individuals.

References

LESTHAEGHE R.J. *Second Demographic Transition*, Basil Blackwell. http://www.vub.ac.be/SOCO/ron/final_textSDTBasilBlackwellEncyclop.doc (access on 09/07/2014).

[4] Except Norway, not included due to few observations.

LESTHAEGHE R.J. 2010. The Unfolding Story of the Second Demographic Transition, *Population and Development Review* 36(2): 211-251.

LESTHAEGHE R.J., NEIDERT L. 2006. The Second Demographic Transition in the United States: Exception or Textbook Example? *Population and Development Review* 32(4): 669-698.

LESTHAEGHE, R.J, VAN DE KAA, D.J. 1986. Twee demografische transities. In: Lesthaeghe, R.J and van de Kaa, D.J., (a cura di), *Groei of krimp?*, book edition Mens en Maatshappij, Van Loghum Slaterus, Deventer.

SOBOTKA T. 2004. *Postponement of Childbearing and Low Fertility in Europe*. Amsterdam: Dutch University Press.

SOBOTKA T. 2008. Does persistent low fertility threaten the future of European populations? In J. Surkin at al. (eds.) *Demographic Challenges for the 21st Century*. Brussels: VUBPress, pp.27-90.

SUTTON P., MATHEWS T.J. 2004. Trends in characteristics of births by state: United States, 1990,1995, and 2000-2002. *National Vital Statistics Report*, 52(9). National Vital Statistics System and Centers for Disease Control and Prevention, US Department of Health and Human Services.

VAN DE KAA, D.J. 1987. Europe's Second Demographic Transition. *Population Bulletin*, 42(1).

VAN DE KAA, D.J. 2002. The idea of a second demographic transition in industrialized countries. Paper 6th Welfare Policy Seminar, National Institute of Population and Social Security, Tokyo, 29 January.

ZAGAGLIA B. 2006. I comportamenti riproduttivi nelle società post-transizionali: un puzzle di teorie. Un tentativo di sistemazione teorica. Quaderno n.255, *Quaderni di Ricerca*. Ancona: Dipartimento di Economia, Università Politecnica delle Marche.

SUMMARY

Fertility dynamics in Europe: Reflections on the principal interpretative paradigms in light of some empirical evidence

In this paper, we discuss the principal interpretative paradigms of recent reproductive dynamics in Europe. We focus on the Second Demographic Transition theory and analyze the long-term relationships between childbearing postponement and change in fertility intensities. We find results at odds with the theory.

Barbara ZAGAGLIA, Università Politecnica delle Marche, b.zagaglia@univpm.it
Eros MORETTI, Università Politecnica delle Marche, e.moretti@univpm.it

Rivista Italiana di Economia Demografia e Statistica Volume LXVIII n. 3/4 Luglio-Dicembre 2014

EFFETTI DELL'INVECCHIAMENTO DELLA POPOLAZIONE SULLA SPESA DEL SISTEMA SANITARIO NAZIONALE

Anna Maria Altavilla, Angelo Mazza, Luisa Monaco

1. Introduzione

L'attuale dinamica della popolazione suscita delle preoccupazioni sulla futura sostenibilità del sistema sanitario nazionale in Italia. La crescente quota di soggetti anziani, determinata dal progressivo invecchiamento della popolazione, potrebbe portare nel prossimo futuro a una più elevata incidenza delle patologie cronico-degenerative, a una maggiore richiesta di cure sanitarie e assistenziali e, di conseguenza, a un aumento della spesa sanitaria. Il timore per l'eccessiva espansione della spesa sanitaria è suffragato dai dati rilevati nel recente passato, che mostrano, per i consumi e per la spesa sanitaria pro-capite, un andamento crescente all'aumentare dell'età della popolazione.

Nel presente studio, partendo dalla relazione che per l'ultimo ventennio ha legato l'andamento della spesa sanitaria pubblica pro-capite con l'invecchiamento demografico, saranno presentate le previsioni sull'andamento della spesa sanitaria per il prossimo futuro. Lo studio sarà effettuato per l'intero Paese e sarà esteso a ogni regione del territorio italiano.

2. I consumi sanitari in Italia

Le evidenze empiriche mostrano una correlazione positiva tra consumi sanitari pro-capite e composizione per fasce di età della popolazione. Agli individui anziani si associa, infatti, una maggiore domanda di servizi di assistenza sanitaria e sociosanitaria e, di conseguenza, una più elevata spesa sanitaria pro-capite (Rebba, 2005; Dortmond et al., 2006; Protonotari et al., 2007). Con riguardo alla spesa farmaceutica si registra un fenomeno analogo. I consumi e i costi che interessano la popolazione anziana sono decisamente più alti sia rispetto alla media nazionale, sia rispetto alle fasce di età inferiore (AIFA, 2013).

Dall'analisi della domanda sanitaria complessiva distinta per fasce di età, nonché dei relativi costi, emerge un andamento a "J". Nello specifico, si rileva un

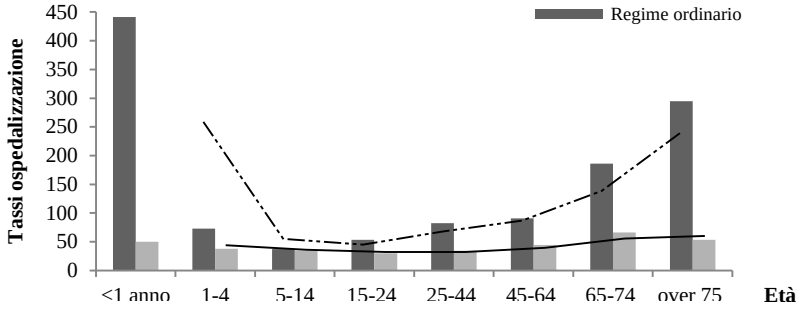
massino locale nella spesa sanitaria riferita ai primi anni di vita degli individui, cui segue un progressivo decremento negli anni successivi all'infanzia, e una ripresa della crescita intorno ai cinquant'anni di età, fino al raggiungimento del livello massimo assoluto nella classe di età 75-80 anni (Rebba, 2005; Gabriele e Raitano, 2009). In letteratura si sottolinea, peraltro, come tra i *driver* della spesa per l'assistenza socio-sanitaria non rientri l'età *per se*, quanto lo stato di salute e le abilità funzionali degli individui anziani (EPC, 2006). Tra la popolazione anziana vi è, infatti, una più elevata incidenza di patologie cronico-degenerative rispetto agli individui più giovani, e da ciò deriverebbe una maggiore pressione sulla spesa sanitaria (EPC, 2006).

In quel che segue, è rappresentato, per l'anno 2012 (ultimo dato disponibile), l'andamento per fasce di età di due importanti voci della domanda sanitaria, ossia i ricoveri ospedalieri e i farmaci. Per i ricoveri sono considerati i tassi di ospedalizzazione riguardanti gli acuti e la riabilitazione, sia in regime ordinario sia in Day Hospital, e la lungodegenza (figg. 1-3). Il consumo farmaceutico (fig. 4), è espresso come numero medio di dosi di farmaco consumate giornalmente ogni 1.000 abitanti (DDD/1.000 abitanti die).

Dai grafici appare evidente una maggiore incidenza dei consumi sanitari nelle fasce di età 65-74 anni e *over* 75. Peraltro, osservando la distribuzione dei consumi sanitari per fasce di età, per il decennio 2002-2012, si rileva un *pattern* pressoché simile (cfr. rapporti Ministero della salute e Aifa, anni 2002-2012).

L'ulteriore invecchiamento della popolazione potrebbe, pertanto, rappresentare un importante fattore di crescita della spesa. In quel che segue, sono presentate le proiezioni sull'andamento della spesa sanitaria pubblica pro-capite al 2025, considerando l'effetto determinato dalla variazione degli indicatori demografici d'invecchiamento, per i quali l'Istat fornisce le previsioni a medio-lungo termine[1].

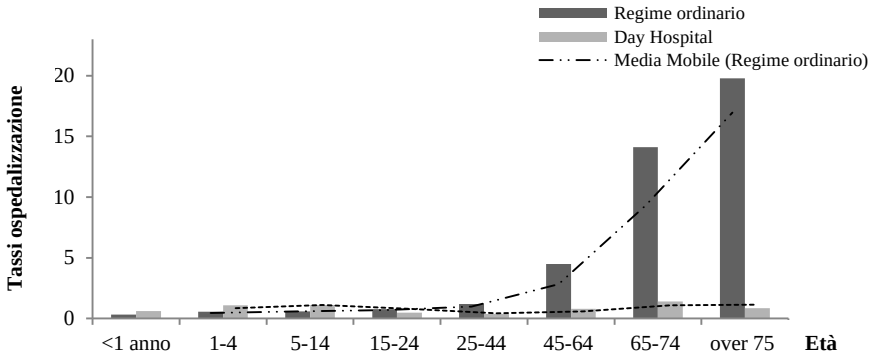
Figura 1 – *Acuti, tassi di ospedalizzazione per fasce di età (per 1.000 abitanti), anno 2012*



Fonte: nostra elaborazione su dati SDO (Ministero della salute, 2012).

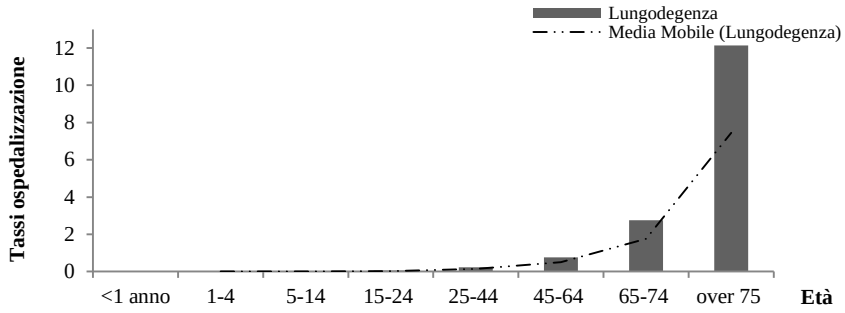
[1] Le proiezioni impiegate nella presente indagine si riferiscono allo "scenario centrale" Istat 2011-2065

Figura 2 – *Riabilitazione, tassi ospedalizzazione per fasce di età (per 1.000 ab.), anno 2012.*



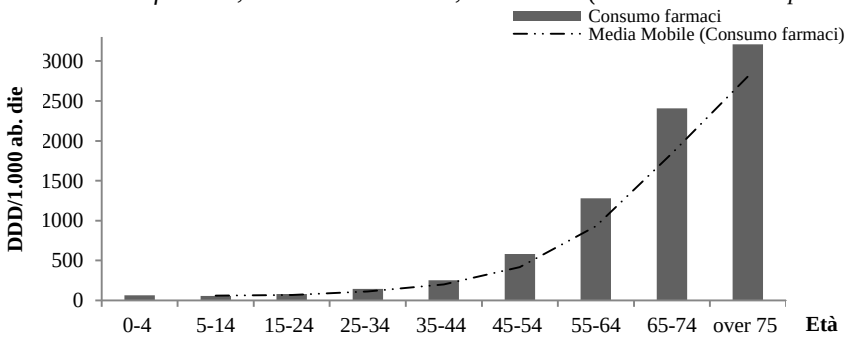
Fonte: nostra elaborazione su dati SDO (Ministero della salute, 2012).

Figura 3 – *Lungodegenza, tassi ospedalizzazione per fasce di età (per 1.000 ab.), anno 2012.*



Fonte: nostra elaborazione su dati SDO (Ministero della salute, 2012).

Figura 4 – *Consumo farmaci, DDD/1.000 ab. die, anno 2012 (esclusi i consumi ospedalieri).*



Fonte: nostra elaborazione su dati AIFA (2013).

3. Proiezioni sull'andamento della spesa SSN pro-capite

3.1. Andamento spesa SSN e dinamica demografica 1990-2011

L'analisi della spesa pro-capite[2] per il servizio sanitario nazionale (SSN), per il periodo 1990-2011[3], è effettuata utilizzando i dati presenti nel database "Health for All" (HfA) fornito dall'Istat. L'esame dell'andamento della spesa in Italia evidenzia un *trend* crescente. Considerando, tuttavia, il dato regionale (Tabella 1) si nota una marcata disomogeneità nel tasso di crescita della spesa. Infatti, l'incremento più basso è riferito alle Marche (+18.3%), mentre il più elevato riguarda la Basilicata (+78.6%), con una differenza tra il valore minimo e massimo di oltre 60 punti percentuali. I fattori di crescita della spesa andrebbero ricercati nella politica sociale, nel progresso tecnologico, in una maggiore consapevolezza della popolazione sul proprio stato di salute cui si accompagna spesso una crescente domanda di prestazioni sanitarie, ma anche nella mutata composizione per età della popolazione. Quanto a quest'ultimo aspetto, ragionevolmente, la progressiva crescita della quota di anziani sulla popolazione totale sarà accompagnata da una maggiore incidenza di patologie, soprattutto cronico-degenerative e, di conseguenza, da un più diffuso ricorso alle cure e all'assistenza socio-sanitaria (Gabriele e Raitano, 2009).

Tabella 1 - *Incrementi % spesa SSN pro-capite, per regioni (1990-2011).*

Regioni	Incrementi %	Regioni	Incrementi %
Marche	18.30	Lombardia	42.73
Emilia-Romagna	19.70	Puglia	46.86
Liguria	26.25	Friuli-Venezia Giulia	46.97
Veneto	26.48	Piemonte	49.63
Toscana	34.01	Trentino-Alto Adige	54.94
Umbria	35.43	Calabria	60.68
Sicilia	40.68	Sardegna	61.52
Campania	41.09	Valle d'Aosta	64.02
Abruzzo	41.76	Molise	66.93
Lazio	42.17	Basilicata	78.62
		ITALIA	39.80

Fonte: nostra elaborazione su dati HfA Italia.

[2] L'aggregato comprende le seguenti voci di spesa pro-capite: per servizi forniti direttamente, in convenzione totale per prestazioni sociali, in convenzione per assistenza farmaceutica, in convenzione per assistenza medico generica, in convenzione per assistenza medico-specialistica, in convenzione per case di cura private, in convenzione per assistenza protesica e cure balneo-termali, per altre prestazioni, per altre spese (HfA).

[3] Ai fini dell'indagine di cui al presente lavoro i dati sono stati rivalutati al 2012.

Va rilevato (Tabella 2) come, in Italia, la percentuale d'individui in età 65+ sia passata dal 14.9% (anno 1990) al 20.6% (anno 2011); al contempo, la quota di *over* 85 (c.d. grandi anziani) è passata dall'1.2% (1990) al 2.8% (2011). Allo stesso modo, gli altri indicatori che misurano l'invecchiamento demografico, quali l'indice di vecchiaia, l'indice di dipendenza anziani, la speranza di vita alla nascita e all'età di 65 anni presentano, dal 1990 al 2011, valori crescenti[4].

Tabella 2 – *Indicatori di invecchiamento in Italia, variazioni annue (1990-2011).*

	Popolaz. residente 65+ (%)	Popolaz. residente 85+ (%)	Indice di vecchiaia	Indice dipendenza anziani	$e_{0,M}$	$e_{0,F}$	$e_{+65,M}$	$e_{+65,F}$
1990	14.90	1.20	90.01	21.73	73.61	80.06	14.99	18.75
2011	20.56	2.82	146.52	31.43	79.40	84.50	18.40	21.90
Δ 1990-2011	0.27†	0.08†	2.69†	0.46†	0.37*	0.26*	1.08*	0.80*

†*punti percentuali * variazione percentuale*
Fonte: nostre elaborazioni su dati ISTAT

3.2. Proiezioni al 2025

Per stimare il futuro andamento della spesa SSN sarebbe utile costruire un modello che valuti il contributo di tutti i fattori di crescita della spesa. Tale tipo di analisi, tuttavia, trova un limite nella difficoltà di individuare e quantificare il contributo apportato da ciascun fattore. Per tale motivo, si è scelto di focalizzare l'analisi sull'invecchiamento demografico; rispetto a quest'ultima variabile, infatti, si dispone di un ampio *set* informativo, sia con riferimento all'estensione temporale dei dati (osservati e previsti) sia con riguardo al numero di indicatori che misurano il fenomeno.

Tabella 3 – *Indicatori d'invecchiamento in Italia, variazioni annue (2012-2025).*

	Popolaz. residente 65+ (%)	Popolaz. residente 85+ (%)	Indice di vecchiaia	Indice dipendenza anziani	$e_{0,M}$	$e_{0,F}$	$e_{+65,M}$	$e_{+65,F}$
2012	20.60	2.90	146.80	31.50	79.70	84.80	18.50	22.10
2025	23.90	4.20	183.60	37.90	82.00	87.00	20.10	3.90
Δ 2012-2025	0.25†	0.10†	2.83†	0.49†	0.22*	0.20*	0.67*	0.63*

†*punti percentuali * variazione percentuale*
Fonte: nostre elaborazioni su dati ISTAT

[4] Ciò ha prodotto un effetto sull'età media della popolazione che, dal 1990 al 2011, è passata da 39.7 a 44.9 anni.

In una prima fase dell'indagine sono stati calcolati i coefficienti di correlazione tra la spesa sanitaria e gli indicatori d'invecchiamento. Sono stati poi sviluppati due differenti metodi di stima: una regressione lineare OLS e una regressione non parametrica, impiegando lo stimatore LOWESS robusto (Cleveland, 1979).

Tabella 4 – *Incrementi % medi annui spesa SSN pro-capite (valori osservati 1990-2011, valori previsti 2012-2025)*

Regioni	Valori osservati	Valori stimati	
		Regressione lineare	Regressione non parametrica
	1990-2011	2012-2025	2012-2025
Piemonte	2.36	1.51	0.99
Valle d'Aosta	3.05	1.84	1.60
Lombardia	2.04	1.32	1.06
Trentino-Alto Adige	2.62	1.20	0.02
Veneto	1.26	1.13	0.90
Friuli-Venezia Giulia	2.24	1.95	1.51
Liguria	1.25	0.90	0.77
Emilia-Romagna	0.94	0.97	0.50
Toscana	1.62	1.14	0.99
Umbria	1.69	1.06	0.62
Marche	0.87	0.86	1.11
Lazio	2.01	1.39	-0.93
Abruzzo	1.99	1.62	0.16
Molise	3.19	1.85	1.45
Campania	1.96	1.76	0.06
Puglia	2.23	1.53	1.43
Basilicata	3.74	1.59	1.60
Calabria	2.89	1.37	1.17
Sicilia	1.94	1.60	0.46
Sardegna	2.93	1.28	1.10
ITALIA	1.90	1.39	0.01

Fonte: nostre elaborazioni su dati ISTAT

In entrambi i metodi, la variabile dipendente è data dalla spesa SSN pro-capite osservata per il periodo 1990-2011 e rivalutata al 2012; la variabile esplicativa è stata scelta selezionando l'indicatore maggiormente correlato con la spesa sanitaria[5]. In Tabella 3 sono riportati gli indicatori impiegati nelle stime e il relativo incremento medio annuo (per l'Italia) per il periodo 2012-2025. Dal confronto

[5] Va rilevato, a tale proposito, che l'analisi è stata effettuata distintamente per tutte le regioni e che l'indicatore più strettamente correlato con la spesa SSN è risultato, in prevalenza, la speranza di vita alla nascita o a 65 anni.

delle variazioni annue osservate (Tabella 2) con quelle previste (Tabella 3) si nota come per queste ultime l'incremento sia, in generale, più attenuato che nel passato.

Le proiezioni sulla spesa SSN pro-capite al 2025 sono riferite alle singole regioni italiane e all'Italia. Nella tabella 4 sono riportate, per ciascuna regione e per l'Italia, le stime dell'incremento che si potrebbe osservare, per la spesa sanitaria al 2025, per effetto del progressivo invecchiamento demografico. L'esame dei dati evidenzia come le stime basate sul modello non parametrico appaiano più contenute rispetto alla regressione lineare. Inoltre, l'R-quadro denota un migliore adattamento dei dati per la stima non parametrica.

Confrontando i risultati ottenuti con i due criteri di stima, in entrambi i modelli si rileva, per quasi tutte le regioni italiane, una variazione media annua positiva della spesa, tuttavia inferiore rispetto al passato. La regione Marche presenta un incremento medio della spesa stimata superiore rispetto al passato. Per il Lazio, le previsioni danno ragione di un possibile decremento medio annuo, peraltro in linea con l'andamento della spesa osservata negli ultimi anni del passato decennio.

4. Conclusioni

Nel presente lavoro sono state presentate le previsioni sull'andamento della spesa SSN pro-capite al 2025, con riferimento alle regioni italiane. Dalle stime, condotte sia con un modello di regressione lineare, sia con una regressione non parametrica, si evidenzia per il prossimo futuro una crescita della spesa. L'incremento, per quanto più contenuto rispetto al passato, desta delle preoccupazioni sull'effettiva sostenibilità dei costi legati alla sanità pubblica. Considerata l'impossibilità, almeno nel breve periodo, di controllare la dinamica demografica, il contenimento della spesa SSN potrebbe essere realizzato in via indiretta, attraverso interventi di *policy* finalizzati a migliorare lo stato di salute degli individui anziani, tra i quali vi è una maggiore incidenza di patologie cronico-degenerative. Rientrerebbero tra questi le azioni destinate alla popolazione in età giovane e media, riguardanti l'ambito della prevenzione e della diagnosi precoce.

Riferimenti bibliografici

AIFA-AGENZIA ITALIANA DEL FARMACO. 2013. L'uso dei Farmaci in Italia. Rapporto Nazionale Anno 2012, Rapporto dell'Osservatorio Nazionale sull'Impiego dei Medicinali (OSMED). Roma.

CLEVELAND, W. 1979. Robust locally weighted regression and smoothing scatterplots. Journal of the American statistical association. Vol. 74. No. 368. pp. 829-836.

DORMONT, B., GRIGNON, M., HUBER, H. 2006. Health expenditure growth: reassessing the threat of ageing. Health economics. Vol. 15. No. 9. pp. 947-963.

GABRIELE, S., RAITANO, M. 2009. Invecchiamento, salute, spesa sanitaria e di cura in Italia. Studi e note di economia. Vol. 14. No. 3. pp. 541-569.

EPC-ECONOMIC POLICY COMMITTEE. 2006. Impact of Ageing Populations on Public Spending on Pensions, Health and Long-Term Care, Education and Unemployment Benefits for the Elderly: Summary Report. Brussels.

MINISTERO DELLA SALUTE. 2013. Rapporto annuale sull'attività di ricovero ospedaliero. Dati DSO 2012. Roma.

PROTONOTARI, A., BURIANI, A., MEZZETTI, F., ET AL. 2007. Analisi economica sull'assorbimento di risorse da parte della popolazione anziana dell'Ausl di Bologna. Politiche sanitarie. Vol. 8. pp. 89-94.

REBBA, V. 2005. Invecchiamento della popolazione e spesa per il long term care. Politiche sanitarie. Vol. 6. No. 3. pp. 109-115.

SUMMARY

Effects of ageing population on public health expenditure

In this paper, we provide a forecast of the trend in the Italian public health expenditure. The analysis is based on the relation between *per capita* health expenditure and population ageing, occurred over the last twenty years; both a linear regression model, as well as a robust locally weighted regression (LOWELL) model were fitted. In both methods, as regressors we employed ageing indexes, selected among those provided by Istat. Results up to the year 2025 mainly confirm the rising trend in the public health expenditure.

Anna Maria ALTAVILLA. University of Catania. Department of Economics and Business. altavil@unict.it

Angelo MAZZA. University of Catania. Department of Economics and Business. a.mazza@unict.it

Luisa MONACO. University of Catania. Department of Economics and Business. luisa.monaco@unict.it

Rivista Italiana di Economia Demografia e Statistica Volume LXVIII n. 3/4 Luglio-Dicembre 2014

A COMPARISON OF BIAS CORRECTION METHODS FOR THE DISSIMILARITY INDEX

Anna Maria Altavilla, Angelo Mazza, Antonio Punzo

1. Introduction

The segregation of demographic groups, often connected to ethnicity, age or gender, is an important area of research among sociologists, demographers and other social scientists. The evaluation of segregation within a population is typically based on the proportions of demographic groups belonging to some kind of allocation units, such as residential areas, workplaces, or schools (Mazza and Punzo, in press).

Many segregation indexes have been suggested, with different formulations denoting different definitions of segregation (see Massey and Denton, 1988 for an overview). Among these, the dissimilarity index *D*, proposed by Duncan and Duncan (1955), is widely used to assess the differential distribution of two groups among allocation units. This index has been used in a broad range of contexts, such as gender segregation (see, e.g., Karmel and Maclachlan, 1988), labor force segregation (for a survey see Flückiger and Silber, 1999), and residential segregation (see Duncan and Duncan, 1955, and Massey and Denton, 1987, 1988).

Generally, the observed settlement pattern is the resultant of a mix of behavior-based forces; thus it should be seen as one of the many possible outcomes of a stochastic - rather than deterministic - allocation. Usually researchers are interested in understanding the "systematic" characteristics of the allocation process, apart from random fluctuations that may affect a single observed pattern (Altavilla, Mazza, Punzo, 2012). In this view, the observed dissimilarity $\widehat{D}$ is merely an estimator of a true but unknown level of dissimilarity in the population D. So, it should be clear why this randomness also holds even if the index is computed on a full-count census data. A problem with the use of this index is that $\widehat{D}$ appears to be an upward biased estimator of D. Within a multinomial framework based on the assumption that individuals allocate themselves independently and that unit sizes are not fixed (see Section 2), Allen et al. (2009) demonstrate, using simulations, that random allocation generates substantial unevenness, and hence an upward bias, especially when dealing with small units, a small minority proportion, and a low level of seg-

regation. Accordingly, different correction approaches have been proposed in literature (see, e.g., Allen et al., 2009, and Altavilla, Mazza and Punzo, 2010 for two examples of bootstrap-based bias correction, Altavilla, Mazza, Punzo, 2012 for an analytical computation of bias and Mazza and Punzo 2014 for a new bias correction which outperforms all previous correction attempts).

In the following, four bias correction techniques, based on grouped jackknife, bootstrap, double bootstrap and the Mazza and Punzo (in press) proposal, are compared in terms of their mean bias. The paper is organized as follows. In section 2, inferential framework and notation are given, in section 3 the four estimators are described and in section 4 there is their comparison. Finally, in section 5, conclusions are drawn.

2. Inferential framework and notation

Consider an area subdivided into k subareas (or units), denoted by $j = 1, \ldots, k$, being populated by n individuals according to a dichotomous characteristic indexed by $c = 0,1$. Examples of common dichotomous characteristics are black or white ethnicity, male or female gender, and so on. The number of individuals with status c is denoted by n^c, $c = 0,1$, with $n = n^0 + n^1$. There will be n_j^c individuals in unit j having status c, with $n^c = \sum_{j=1}^{k} n_j^c$, $c = 0,1$. The observed settlement -- characterized by the two sets denoted by $n_1^0, \ldots, n_k^0$ and $n_1^1, \ldots, n_k^1$ -- is, however, only one of the possible realizations of an underlying *allocation process* P. If it is plausible to assume that individuals allocate themselves independently and that unit sizes are not fixed, then the process will be governed by the conditional probabilities

$$p_j^c = P(unit\ of\ membership = j|c), \qquad j = 1, \ldots, k, \qquad c = 0,1 \qquad (1)$$

that an individual i will belong to the unit j, given his/her status c.

Social scientists are usually interested in making inferences on a particular function of these probabilities; this function, commonly called "segregation index", should express the degree of segregation that characterize the process P. Before to introduce any kind of segregation index, it is important to define the concept of systematic segregation, occurring when there is at least one subarea in which individuals belonging to the two groups have a different probability to allocate themselves; in mathematical terms this means that:

$\exists\ j:\ p_j^1 \neq p_j^0.$

Among the many segregation indexes existing in literaturem, the most popular one is without doubt the Duncan and Duncan (1955) segregation index, usually denoted by D, characterized by the formula:

$$D = \frac{1}{2}\sum_{j=1}^{k} \left| p_j^1 - p_j^0 \right| \tag{2}$$

Obviously, the index in (2) takes values on the compact interval $[0,1]$ and it increases as systematic segregation grows. Furthermore, it is straightforward to note that the case $D = 0$ (absence of systematic segregation) is achievable if, and only if

$p_j^1 = p_j^0 \quad \forall\, j$.

Unfortunately, we can only observe the crude counterpart of D

$$\widehat{D} = \frac{1}{2}\sum_{j=1}^{k} \left| \frac{N_j^1}{n^1} - \frac{N_j^0}{n^0} \right| = \frac{1}{2}\sum_{j=1}^{k} \left| \hat{p}_j^1 - \hat{p}_j^0 \right| \tag{3}$$

where $\hat{p}_j^c$, proportion of individuals with status c in the unit j, $c = 0,1$, is the plug-in estimator of p_j^c. The word "unfortunately" is justified if one thinks that the observed settlement pattern is only one of the numerous possible patterns arising from P, each of them with probability (see Allen *et al*., 2009) given by the product of two independent multinomial distributions, one for $c = 0$ and one for $c = 1$:

$$P(n_1^c, \ldots, n_k^c \mid p_1^c, \ldots, p_k^c, n^c) = \prod_{j=1}^{k} \prod_{c=0}^{1} n^c! \frac{\left(p_j^c\right)^{n_j^c}}{n_j^c!} \tag{4}$$

3. Estimators

In this section, we introduce four alternative bias correction techniques.

1.1. Bootstrap based estimator

With the aim to eliminate, or at least reduce, the upward bias of $\widehat{D}$, Allen et al. (2009) adopt a bootstrap-based bias correction. It is based on the idea that

$$D - \widehat{D}_{\text{obs}} \approx \widehat{D}_{\text{obs}} - E\left(\widehat{D}|\hat{p}_1^0, \dots, \hat{p}_k^0, \hat{p}_1^1, \dots, \hat{p}_k^1, n^0, n^1\right), \quad (5)$$

where $\widehat{D}_{\text{obs}}$ denotes the observed counterpart of $\widehat{D}$. The observed conditional probabilities $\hat{p}_j^0$ and $\hat{p}_j^1$, $j = 1, \dots, k$, are used to generate, by multinomial sampling, B bootstrap allocations with the same group sizes n^0 and n^1. Then, a measure of $\text{Bias}\left(\widehat{D}\right)$ is given by $\overline{D}_{\text{Boot}} - \widehat{D}$, and the bootstrap bias corrected estimate of D can be obtained as

$$\widehat{D}_{\text{Boot}} = \widehat{D}_{\text{obs}} - \left(\overline{D}_{\text{Boot}} - \widehat{D}_{\text{obs}}\right) = 2\widehat{D}_{\text{obs}} - \overline{D}_{\text{Boot}}. \quad (6)$$

This bias correction would work well if the bias were constant for different values of D. This is not the case here, and this bias correction is therefore not expected to "eliminate", but only to "reduce", the existing bias. Instead of bootstrapping $E\left(\widehat{D}|\hat{p}_1^0, \dots, \hat{p}_k^0, \hat{p}_1^1, \dots, \hat{p}_k^1, n^0, n^1\right)$, Mazza and Punzo (in press) show that this expectation may be computed analytically, using a binomial based formulation for a small number of units with small sizes or with a folded normal approximation when n^c, $c = 0{,}1$, is sufficiently large.

1.2. Grouped jackknife and iterative bootstrap estimators

Alternative to the bootstrap, a standard practice for bias correction is the Jackknife. Hence, we evaluated, also, a grouped jackknife estimator $\widehat{D}_{\text{JK}}$; this estimator has been implemented following Efron (1982, Section. 2.2). Finally, a double bootstrap estimator $\widehat{D}_{\text{DB}}$, based on the approach documented in Davison and Hinkley (1997, Section. 3.9) has also been evaluated.

1.3. A recently introduced estimator

Mazza and Punzo (2014) introduce a new estimator of D, which further reduces the bias with respect to $\widehat{D}_{\text{Boot}}$. Its rationale consists in choosing a value $\widetilde{D}$ which minimizes

$$E\left(\widehat{D}|\tilde{p}_1^0, \dots, \tilde{p}_k^0, \tilde{p}_1^1, \dots, \tilde{p}_k^1, n^0, n^1\right) - \widehat{D}_{\text{obs}} \quad (7)$$

with $\widetilde{D} = \frac{1}{2}\sum_{j=1}^{k} \left|\tilde{p}_j^1 - \tilde{p}_j^0\right|$. There may be different criteria for choosing $\widetilde{D}$. On way is to require the sequence of differences $\left|\tilde{p}_j^0 - \tilde{p}_j^1\right|$ to be a flattened variant of its observed counterpart. Flattening is obtained by spreading the difference $\Delta = \widehat{D}_{\text{obs}} - \widetilde{D} \geq 0$, among the k differences $\left|\tilde{p}_j^0 - \tilde{p}_j^1\right|$, proportionally to the residuals $\hat{d}_j = \left|\hat{p}_j^0 - \hat{p}_j^1\right|$. An optimization procedure, which adopts a combination of golden section search and successive parabolic interpolation is described in Mazza and Punzo (2014).

4. Comparison of estimators

In this section we use Monte Carlo simulations to compare the bias of $\widehat{D}$, and of the four estimators $\widehat{D}_{\text{Boot}}$, $\widehat{D}_{\text{JK}}$, $\widehat{D}_{\text{DB}}$ and , $\widetilde{D}$ described in the previous section. The setup of the simulations is similar to the one adopted by Allen et al. (2009). The sets of conditional probabilities $p_1^0,\ldots,p_k^0$ and $p_1^1,\ldots,p_k^1$, with $k=50$, were obtained with the formula

$$P(\text{unit} \le j \mid c=1) = \frac{(1-q)P(\text{unit} \le j \mid c=0)}{1-qP(\text{unit} \le j \mid c=0)} \tag{8}$$

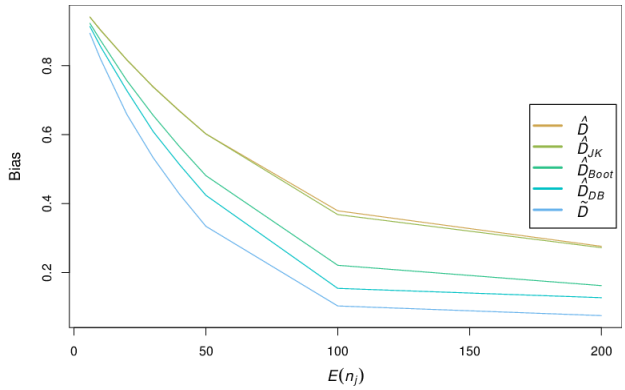
proposed in Duncan and Duncan (1955); it may be observed that each value of q is related to one value of D. Although this set of segregation curves cannot represent all distributions of segregation, it is a sufficient set to examine different levels of systematic segregation for the purposes of this paper. The formula above, combined with the constraint of equal expected unit sizes $E(n_j)$, fixes the conditional allocation probabilities for both groups. An allocation is then generated by assigning n^1 and n^0 individuals to the k units by sampling from two multinomial distributions having each one of the two sets of conditional probabilities as parameter.

The simulation factors considered are p, $E(n_j)$ and D. For each of them, a grid of values is chosen: 0.01, 0.05, 0.1, 0.3, and 0.5 for p; 6, 10, 20, 30, 40, 50, 100 and 200 for $E(n_j)$; 0, 0.056, 0.127, 0.225, 0.292, 0.382, 0.634, and 0.818 for D. Values chosen for D are respectively related to the values 0, 0.2, 0.4, 0.6, 0.7, 0.8, 0.95, and 0.99, of the parameter q in the previous equation. The number of units is fixed at $k = 50$ and the number of bootstrap replications is fixed to B = 100. For each combination of the considered simulation factors, 1000 samples are generated randomly.

The mean simulated biases of the estimators considered are depicted in the figures below.

It may be noted that when p, $E(n_j)$ and D present low values, the bias of $\widehat{D}$, the uncorrected estimator, is considerably high, incorrectly suggesting that a highly segregating process underlies the allocation. In the opposite situation of high values of p, $E(n_j)$ and D, all estimators provide values not very different from the true value D. From these results, we can note as $\widetilde{D}$ most often outperforms all other estimators in reducing the bias, while the grouped jackknife estimator, in all the considered scenarios of simulations, showed only a negligible improvement over $\widetilde{D}$.

Figure 1 – *Comparison between biases at the varying of* $\boldsymbol{E(n_j)}$*, fixed p = 0.01 and* $\boldsymbol{D}$*=0 .*



As to the double bootstrap approach, the added level of bootstrap did improve the performance of in terms of mean bias over $\widehat{D}_{\mathrm{Boot}}$; however, these improvements were only marginal, and very far from counterbalancing the higher computational burden required.

Figure 2 – *Comparison between biases at the varying of D, fixed p = 0.01 and* $\boldsymbol{E(n_j)}$ *= 20.*

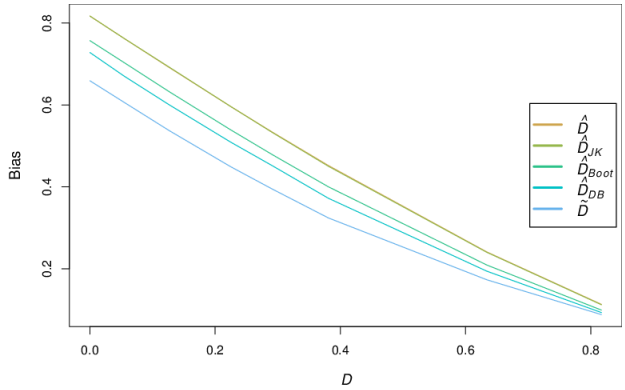
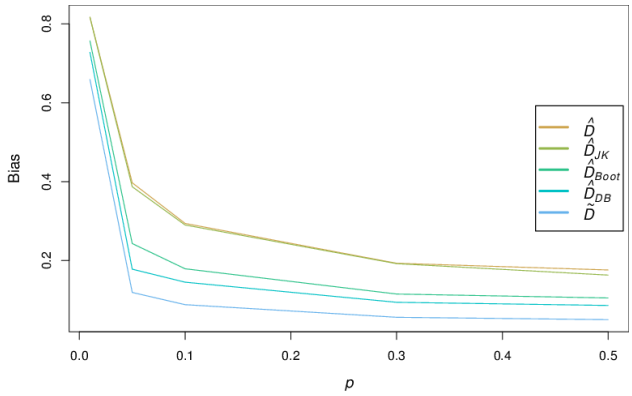


Figure 3 – *Comparison between biases at the varying of p, fixed D = 0 and* $\boldsymbol{E(n_j)}$ *= 20.*



5. Conclusions

It has long been recognized that the sensitivity of the dissimilarity index of Duncan and Duncan (1955) to random allocation implies an upward bias, particularly evident with smaller unit sizes, small minority proportions and lower levels of segregation. In this paper, following a multinomial framework, we have compared, using Monte Carlo simulations, the performance of four bias reduction techniques, based on bootstrap, grouped jackknife, double bootstrap and on a recent procedure introduced in Mazza and Punzo (2014). This new procedure performed better than its competitors did, although for reliable estimations, minority proportion and unit sizes do not have to be both very small. The grouped jackknife bias-corrected estimator exhibited only a little improvement over the natural estimator and so did the double bootstrap estimator with respect to the bootstrap bias-corrected one.

References

ALLEN, R., BURGESS, S., WINDMEIJER, F. (2009). More reliable inference for segregation indices. *Technical Report 216*, The Centre for Market and Public Organisation, University of Bristol.

ALTAVILLA A.M., MAZZA A., PUNZO A. (2010). Sull'impiego di un indice di dissimilarità nello studio della disposizione di popolazioni straniere su un territorio urbano. *Rivista Italiana di Economia, Demografia e Statistica*, vol. LXIV, p. 7-14.

ALTAVILLA A.M., MAZZA A., PUNZO A. (2012). On the upward bias of the dissimilarity index. *Rivista Italiana di Economia, Demografia e Statistica*, vol. LXVI – N. 1, p. 15-20.

DAVISON, A. C., HINKLEY, D. V. (1997). Bootstrap Methods and Their Application, volume 1 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge.

DUNCAN, O. D., DUNCAN, B. (1955). A methodological analysis of segregation indexes. *American Sociological Review*, 20(2), 210–217.

EFRON, B. (1982). The Jackknife, the Bootstrap, and Other Resampling Plans, volume 38 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania.

FLÜCKIGER, Y., SILBER, J. G. (1999). *The measurement of segregation in the labor force*. Physica-Verlag, Heidelberg.

KARMEL T., MACLACHLAN M. (1988). Occupational sex segregation - increasing or decreasing? *Economic Record*, 64(3), 187–195.

MASSEY D. S., DENTON, N. A. (1988). The dimensions of residential segregation. *Social Forces*, 67(2), 281–315.

MAZZA A., PUNZO A. (in press). On the upward bias of the dissimilarity index and its corrections. *Sociological Methods & Research*.

SUMMARY

The dissimilarity index of Duncan and Duncan is widely used in a broad range of contexts to assess the overall extent of segregation in the allocation of two groups in two or more units. Its sensitivity to random allocation implies an upward bias with respect to the unknown amount of systematic segregation. In this paper, following a multinomial framework based on the assumption that individuals allocate themselves independently and that unit sizes are not fixed, we report the results of Monte Carlo simulations performed in order to compare the natural estimator with four bias reduction techniques, based on bootstrap, grouped jackknife, double bootstrap and on a more recent procedure. Results indicate the new procedure performed better than its competitors did, although for reliable estimations, minority proportion and unit sizes do not have to be both very small.

Anna Maria ALTAVILLA, University of Catania, Department of Economics and Business, altavil@unict.it

Angelo MAZZA, University of Catania, Department of Economics and Business, a.mazza@unict.it

Antonio PUNZO, University of Catania, Department of Economics and Business, antonio.punzo@unict.it

Rivista Italiana di Economia Demografia e Statistica Volume LXVIII n. 3/4 Luglio-Dicembre 2014

FOREIGN DIRECT INVESTMENT AND PSYCHIC DISTANCE: A GRAVITY MODEL APPROACH

Gianni Bergamo, Claudio Pizzi

1. Introduction

In last decades the internationalization and globalization process has been studied from different perspectives. Different analyses, at both macroeconomic and microeconomic levels, allowed to highlight the interconnection between economic areas or to understand the mechanism that drives trade flows, foreign direct investments, technology transfer and so on. Economists, econometricians and statisticians have proposed theories and models in order to interpret these phenomena, among these the gravity model is an interesting one.

In empirical studies, the gravity model has been widely used in the analysis of trade flows between countries. This model takes into account macroeconomic variables and the distance between countries.

Initially the distance had only a geographical connotation as a proxy for transport costs. Afterwards, the meaning of "distance" has been extended considering also psychological and cultural factors, which could be an obstacle to the information flows between markets. Such dimension is called "psychic distance".

In most of the studies the psychic distance has been measured through a composite index (Kogut and Singh, 1988), that combines the Hofstede's cultural indicators (Hofstede, 1980). Taking into account only the cultural dimension, this approach seems to be limited (Dow, 2000). Recently the psychic distance has been broadened by other dimensions.

In this work we analyse FDI flows through a gravity equation, considering the effect of some variables that influence psychic distance.

After an introduction on the application of the gravity model in economics in section 2, we provide a brief description of the concept of the psychic distance and its role in a gravity equation in section 3. In section 4 we present the research questions and the variables considered in our application. In the last section we present results and discussion.

2. The gravity model

A domestic firm that plans to open a business in a foreign country has to take into account both opportunities and risks arising from this venture. In other words, the decision that firm makes can be viewed as the result of attraction and repulsion forces. The dimension of the economy of a foreign country, for instance, might be viewed as an attraction force as it represents the opportunity, for the firm, of a new end market, while the distance from the foreign market might be viewed as a force in opposite direction because it entails growing transport costs.

Broadening the perspective from a micro to a macroeconomic level, it has been empirically observed that the volume of trade between two countries is proportional to the size of their economies and inversely proportional to their distance (Krugman and Obstfeld, 2009). Generally, the higher is the Gross Domestic Product (GDP), the higher are the trade flows; contrariwise, the higher is the distance and thus the transport costs, the smaller are the trade flows. So, distance seems to have a negative impact on trade flows. These empirical observations have been translated into a formal model by mimicking an astronomical law: the Newton's law of universal gravitation which states that two bodies in the universe attract each other with a force following the relation:

$$F_{1,2} = G\frac{M_1 M_2}{D_{1,2}^2} \tag{1}$$

where $F_{1,2}$ is the attraction force between two astronomical bodies, M_1 and M_2 are their masses, $D_{1,2}$ is the distance between the centre of masses and G is the gravitational constant.

In this paper we study foreign direct investments considering a model derived from (1). More specifically we consider the following gravity model:

$$Y_{i,j} = \beta_0 \frac{X_i^{\beta_1} X_j^{\beta_2}}{D_{i,j}^{\beta_3}} \tag{2}$$

where $Y_{i,j}$ is the volume of FDI from origin country i to the destination country j, X_i and X_j are the size of the economy of the two countries and $D_{i,j}$ is the distance between the two countries. Applying a logarithmic transformation, the model assumes the following additive form

$$\log Y_{i,j} = \log\beta_0 + \beta_1 \log X_i + \beta_2 \log X_j + \beta_3 \log D_{i,j} \tag{3}$$

This transformation enables us to estimate easily the parameters of the linear model using for instance OLS estimators.

The gravity model has been introduced by Tinbergen (1962) at the beginning of the sixties, but only at the end of the seventies, with the contribution of Anderson (1979), it reached a first theoretical formalization. For a long time, indeed, the gravity model has been considered able to capture the empirical regularities, but without a consolidated theoretical basis.

The model (3) is the simplest one and we have to define more specifically the independent variables, in particular the distance that might be considered not only in a geometrical strictly sense (think to a geographical distance between two countries) but we might to enrich this idea by considering several other dimensions such as psychological and cultural ones, each of which need an adequate operationalization.

3. The Psychic distance

In the previous section we have pointed out that the gravity models, at the beginning, have been used to model trade flows. For this reason the distance variable initially was essentially the geographical distance between the two countries involved in the bilateral trade, viewed as a proxy of the transport costs. Afterwards the meaning of the distance has become a multidimensional construct of which the geographical is only one of its dimensions.

By analogy in the decision to invest in a foreign country an investor takes into account some costs: the transport cost is as much important as the information related transaction costs for which the psychic distance is an interesting proxy. So cultural and psychological variables could drive the firm to select the end market (Beckerman, 1956, Johanson and Vahlne, 1977).

The concept of psychic distance, introduced by Beckerman (1956), has been afterwards developed by scholars at the University of Uppsala in the seventies (Johanson and Wiedersheim-Paul, 1975, Johanson and Vahlne 1977, Vahlne and Wiedersheim-Paul, 1977). In particular, Johanson (1977) defined it as: "*... the sum of factors preventing the flow of information from and to the market. Examples are differences in language, education, business practices, culture, and industrial development.*"

Among all, we remind the use of the Hofstede's cultural dimensions (Hofstede, 1980) or the Sethi's market similarity factors (Sethi, 1971). Nevertheless according to Dow (2000), these approaches do not measure the entire construct of psychic distance because they lack some of its components.

Another empirical approach has faced the problem of the measure of the psychic distance by asking key informants to estimate it using a Likert Scale (Nordstrom, 1991, Vahlne and Nordstrom, 1992).

4. Method

4.1. Hypotheses

Taking cue from Dow and Karunaratna (2006) and Dow and Ferencikova (2010), that embed what they called "psychic distance stimuli" into a gravity model, we consider this framework to modelize the FDI flows between couples of countries. The psychic distance perceived by a decision-maker is influenced by several stimuli that should influence the individual and collective perception of it and that are measured using several macro-level variables representatives of differences in language, education, religion, industrial development and political system, as well as the difference in the Hofstede's cultural dimensions and in time zones. For convenience, hereafter we will use "psychic distance" or "psychic distance stimuli" indifferently.

Unlike Dow and Ferencikova (2010) which tested psychic distance influence on market selection prediction, entry mode choice and performance, we consider the gravity model to identify the variables that impact on the level of FDI between couples of countries.

In this work we want to verify the following hypothesis:

H1. The greater the GDPs of the countries between which there is an investment the greater the FDI;

H2. The greater the geographical distance the lower the FDI;

H3. The greater the psychic distance the lower the FDI.

4.2. Variables and sample

To verify the hypotheses H1-H3 we considered the model (2) in which the dependent variable is the average of the yearly investment of the country *i* in the country *j* in the period 2007-2011. We denote this variable by $FDI_{i,j}$.

As far as the independent variables are concerned, in addition to the base-gravity model variables we embed those representing additional barriers or facilitations to the information flows.

To be more precise the economic size of each country is captured by the average GDP from 2007 to 2011 and the psychic distance is captured following Dow and Karunaratna (2006), although some indicators have been recalculated or

replaced by new ones. Specifically, the indicators concerning language (*Lang*) and religion (Relig) are the same made available by Dow and Karunaratna (2006) (see Dow and Karunaratna (2006) and details therein).

Other indicators considered by Dow and Karunaratna (2006), for the operationalization of the psychic distance, may change over time, especially in developing countries[1]. For this reason we updated the variables differences in political system (*Pol*) and in ideology (*Soc*), while we replaced the variable that measures the difference in industrialization with the difference in *Competitive Industrial Performance Index*[2] (CIP).

As Hofstede's cultural dimension (*Hof*) are expected to change over time (Taras *et al.*, 2012) so we used its latest version (Hofstede *et al.*, 2010). A composite index was calculated through an improved version of Kogut and Singh's measure proposed by Kandogan (2012) that takes into account the correlations between the cultural dimensions.

As in Dow and Karunaratna (2006), the difference in time zones (*DTZ*) is measured by the residuals of a linear regression between differences in time zones and geographical distances, in order to remove the dependence of the first variable upon the second.

In our analysis we consider 54 countries[3] overall, half of which are OECD members. The empirical analysis has been restricted to this subset of countries due to data availability on psychic distance stimuli from Dow and Karunaratna (2006). Furthermore, the dependent variable of the model has been extracted from the OECD database, which includes only outward FDI relative to member countries towards partners, whether OECD members or not. Thus, it is not possible to use simmetrically all the feasible country pairs.

The final sample size is 670 observations as we have had to remove some records with null or negative values to avoid undefined logarithms.

[1] For instance, approximately 25 years ago the literacy rate of Turkey was 75.97% and 90.82% in 2009 http://www.quandl.com/society/adult-literacy-rate-all-countries

[2] CIP is provided by UNIDO (http://www.unido.org).

[3] Australia, Bangladesh, Austria, Brasile, Belgio, Cina, Svizzera, Costa Rica, Cile, Ecuador, Spagna, Guatemala, Estonia, Croazia, Finlandia, Indonesia, Francia, India, Grecia, Iran, Ungheria, Lituania, Irlanda, Lettonia, Israele, Marocco, Italia, Malesia, Giappone, Pakistan, Lussemburgo, Panama, Messico, Filippine, Norvegia, Singapore, Nuova Zelanda, El Salvador, Polonia, Suriname, Portogallo, Thailandia, Slovacchia, Vietnam, Slovenia, Russia, Svezia, Venezuela, Turchia, Colombia, Stati Uniti, Perù, Trinidad and Tobago.

5. Results and discussion

The gravity model we use is the multiple linear regression in equation (4)

$$\begin{aligned}\log FDI_{i,j} = \log\beta_0 + \beta_1 \log GDP_i + \beta_2 \log GDP_j + \beta_3 \log D_{i,j} + \\ + \beta_4 \mathrm{Lang_{i,j}} + \beta_5 \mathrm{Relig_{i,j}} + \beta_6 \mathrm{Edu_{i,j}} + \beta_8 \mathrm{Pol_{i,j}} + \\ + \beta_9 Soc_{i,j} + + \beta_{10} Hof_{i,j} + \beta_{12} DTZ.res_{i,j} + \varepsilon_{i,j}.\end{aligned} \quad (4)$$

The β coefficients are estimated through the weighted least squares method (WLS) in order to overcome heteroskedasticity. Table (1) summarize the results.

Table 1 – ***Gravity model summary.***

Variables	Coefficients	Std. Error	Significance
lnGDPi	0.813	0.063	***
lnGDPj	0.979	0.059	***
lnDist	-0.708	0.061	***
Lang	-0.531	0.061	***
Relig	-0.249	0.076	**
Edu	-0.030	0.008	***
CIP	2.106	0.537	***
Pol	-0.071	0.117	
Soc	-0.009	0.047	
Hof	-0.001	0.097	
DTZ	0.001	0.045	
Constant	-38.084	2.010	***
Observations	757		
R2	0.546	Adj. R2	0.539
Resid. Std. Error	1813	(df=745)	
F Statistic	81.463	(df=11; 745)	***
Signif.codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1			

Note: WLS estimators

The coefficients related to the variables *Pol* and *Soc* result not statistically significant. This means that differences in political system and ideological leanings doesn't seem to have an influence on FDI in our sample. Similarly, differences in Hofstede's cultural index (*Hof*) and in time zones (DTZ) are not statistically significant.

The adjusted R^2 is 0.539. A D'agostino K^2 normality test leads to accept the hypothesis of gaussianity of the residuals distribution (p-value 0.95). Collinearity between regressors seems to be absent (variance inflation factors <5).

Hypothesis H.1 is confirmed. Thus, the size of the economy of both countries has a positive effect on the volume of FDI, as expected in a gravity model.

With respect to the hypothesis H.2, as it happens for trade flows, a negative linear effect is present.

Differences in language (*Lang*), religion (*Relig*) and education (*Edu*) have a negative coefficient as expected. The effect of differences in political system and ideology is not statistically significant, as well as the effect of time zones (*DTZ*). The non-significance of the cultural differences (*Hof*) is relevant, since the Hofstede's index is widely used in literature on internationalization. Note that the same results was obtained by Dow and Ferencikova (2010) for the inward FDI market selection prediction, entry mode choice and performance into Slovakia and in Dow and Karunaratna (2006) in a gravity model for trade flows.

In conclusion, the FDI flows between countries are adequately explained by the model here presented, however, further investigation is needed to take into account the different strategies followed by the FDI (horizontal, vertical or conglomerate), and how they are shared among the different countries.

Riferimenti bibliografici

ANDERSON, J., 1979. A theoretical foundation for the gravity equation. The American Economic Review, Vol. 69, No. 1, pp. 106-116.

BECKERMAN, W., 1956. Distance and the pattern of intra-european trade. *Review of Economics and Statistics*, Vol. 38, No. 1, pp. 31-40.

DOW, D., 2000. A note on psychological distance and export market selection. *Journal of International Marketing*, Vol. 8, No. 1, pp. 51–64.

DOW, D. FERENCIKOVA, S., 2010. More than just national cultural distance: Testing new distance scales on FDI in Slovakia. *International Business Review*, Vol. 19, No. 1, pp. 46 – 58.

DOW, D., KARUNARATNA, A., 2006. Developing a multidimensional instrument to measure psychic distance stimuli. *Journal of International Business Studies*, Vol. 37, No. 5, pp. 578–602.

HOFSTEDE, G., 1980. *Culture's Consequences: International Differences in Work-Related Values*. Beverly Hills CA: SAGE Publications.

HOFSTEDE, G., HOFSTEDE, G. J., MINKOV, M., 2010. *Cultures and Organizations: Software of the Mind, Third Edition*. McGraw-Hill Education.

JOHANSON, J., VAHLNE, J.-E., 1977. The internationalization process of the firm - a model of knowledge development and increasing foreign market commitments. *Journal of International Business Studies*, Vol. 8, No. 1, pp. 23–32.

JOHANSON, J., WIEDERSHEIM, P. F., 1975. The internationalization of the firm - four swedish cases. *Journal of Management Studies*, Vol. 12, No. 3, pp. 305–

323.
KANDOGAN, Y., 2012. An improvement to kogut and singh measure of cultural distance considering the relationship among different dimensions of cul-ture. *Research in International Business and Finance*, Vol.26, No.2, pp.196–203.
KOGUT, B., SINGH, H., 1988. The effect of national culture on the choice of entry mode. *Journal of International Business Studies*, Vol. 19, No.3, pp. 411–433.
KRUGMAN, P. R., OBSTFELD, M., 2009. I*nternational Economics: Theory and Policy*. Pearson Addison-Wesley.
SETHI, S. P., 1971. Comparative cluster analysis for world markets. *Journal of Marketing Research*, Vol. 8, No. 3, pp. 348–354.
TARAS, V., STEEL, P., KIRKMAN, B. L., 2012. Improving national cultural indices using a longitudinal meta-analysis of Hofstede's dimensions. *Journal of World Business*, Vol. 47, No. 3, pp. 329–341.
TINBERGEN, J., 1962. *Shaping the World Economy; Suggestions for an International Economic Policy*. Twentieth Century Fund.
VAHLNE, J., WIEDERSHEIM-PAUL, F., 1977. Psychic Distance: An Inhibiting Factor in International Trade. Working paper - CIF. Uppsala University.

SUMMARY

Foreign Direct Investment and Psychic Distance: a Gravity Model Approach

The study of the foreign direct investments (FDI) has, in last decades, captured the interest of several scholars that study, on one hand which variables are crucial in the FDI choices, on the other hand the impact of some variables on the level of FDI between two countries. In this paper we present an empirical analysis, by means of the gravity model, to test the relationship between FDI, GDP, geographical and psychic distance. This distance may be interpreted as an obstacle to the knowledge of foreign markets. In literature, yet few works have explored the effect of psychic distance on FDI, focusing mostly on market selection, entry mode and performance prediction. Furthermore, the use of the Hofstede's index, often used as proxy of psychic distance, it has been shown to be inadequate. To overcome this limit we consider in our application the multidimensionality of the psychic distance by using several variables.

Gianni BERGAMO, Dept of Economics – University Ca' Foscari of Venice, gianni.begamo@unive.it
Claudio PIZZI, Dept of Economics – University Ca' Foscari of Venice, pizzic@unive.it

L'APPROPRIATEZZA DEL TAGLIO CESAREO NELLE REGIONI ITALIANE: ANALISI CON LA CLASSIFICAZIONE DI ROBSON

Francesca Lariccia, Antonella Pinnelli, Sabrina Prati,
Marina Attili, Claudia Iaccarino

1. Introduzione e obiettivi

In Italia il tasso di taglio cesareo (TC) è cresciuto in modo allarmante negli ultimi 30 anni passando dall'11% nel 1980 (Istat, 1980) ad oltre il 36,5% nel 2012 (Ministero della Salute, 2013). Il fenomeno è caratterizzato da una forte variabilità territoriale: le regioni del Nord con tassi generalmente più bassi, e quelle del Sud con tassi elevati. La variabilità territoriale si manifesta anche per tipologia di struttura, con tassi di TC più elevati nei punti nascita privati e in quelli di piccole dimensioni (Ministero della Salute, 2013a).

L'aumento del ricorso al TC non è un fenomeno solo italiano: i parti avvenuti con cesareo sono cresciuti drammaticamente non solo in molti paesi a sviluppo avanzato, ma anche in diversi paesi dell'America Latina e dell'Asia (Betran et al, 2007). Questo è avvenuto nonostante già dal 1985 il WHO raccomandi che il ricorso al TC in ogni paese non dovrebbe superare il 10-15% dei parti (WHO, 1985). In mancanza di prove scientifiche che associno il maggior ricorso al TC ad una diversa distribuzione dei fattori di rischio della popolazione ostetrica o ad un miglioramento degli *outcome* perinatali, queste forti differenze sono un indizio di pratiche cliniche e assistenziali non appropriate (Villar et al, 2006). Dalla letteratura internazionale emerge infatti che la forte crescita che si è osservata in molti paesi è spiegabile solo in piccola parte con indicazioni mediche, mentre un ruolo importante giocano pratiche sanitarie, la preferenza del medico, la richiesta materna (Lavender et al, 2006). In Italia solo dalla fine degli anni '90 è iniziato un intervento istituzionale per arginare e regolare il fenomeno[1], intervento che ad oggi ha avuto scarsa efficacia (Istituto Superiore di Sanità, 2012).

La frequenza dei TC può essere descritta in maniera clinicamente rilevante, standard, analitica e riproducibile nel tempo e nello spazio utilizzando la classificazione di Robson (Robson, 2001). Mediante questo strumento le donne sono classificate in 10 gruppi sulla base di quattro concetti ostetrici e dei relativi

[1] Decreto Ministeriale n. 245 del 20-10-1998 e Decreto Ministeriale 12-12-2001

parametri: categoria della gravidanza, anamnesi ostetrica, andamento di travaglio e parto, età gestazionale. Le 10 classi che si ottengono sono mutuamente esclusive, completamente inclusive, ripetibili nel futuro, e consentono quindi di monitorare e analizzare in un'ottica operativa i tassi di TC tanto a livello locale che globale (Regione Emilia Romagna, 2013; Betran et al, 2014).

Obiettivo di questo lavoro è studiare l'appropriatezza del ricorso al TC nelle regioni italiane utilizzando le classi di Robson. In particolare, si vuole analizzare: quali sottopopolazioni di donne contribuiscono maggiormente agli alti tassi di TC nelle regioni italiane; e se la variabilità tra le regioni nei tassi totali di TC è attribuibile ad una diversa composizione della popolazione assistita, o a differenze nella pratica medico-sanitaria.

2. Materiali e metodi

La costruzione delle classi di Robson e le successive analisi sono state possibili grazie a un database costruito *ad hoc*[2] da F. Lariccia, ottenuto integrando - tramite procedure di Record Linkage (RL) individuale - le informazioni dei Certificati di Assistenza al Parto (CeDAP) relativi al 2003 e dell'Indagine Campionaria sulle Nascite relativa ai nati nello stesso anno (edizione del 2005, l'ultima disponibile). Il database così costruito contiene informazioni socio-demografiche e medico-sanitarie sui nati vivi in Italia e residenti in Italia e comprende 29.089 casi. Il RL ha permesso di validare[3], per la prima volta, le variabili dei CeDAP necessarie alla costruzione delle classi di Robson e quindi di costruire e analizzare le classi per buona parte delle regioni italiane[4].

I tassi di TC regionali sono stati standardizzati per verificare se e in che misura le differenze tra i tassi di TC delle regioni dipendano da una diversa struttura per classi della popolazione ostetrica (e quindi da una differente proporzione di gruppi maggiormente a rischio) o piuttosto da una diversa propensione nella regione a ricorrere al TC (imputabile quindi alle pratiche mediche vigenti nelle strutture, o più in generale, al contesto sanitario). Si è scelto di utilizzare il metodo di

[2] La costruzione del database utilizzato è stata possibile grazie ad uno stage formativo per tesi di dottorato svolto da F. Lariccia presso il Servizio Struttura e Dinamica Demografica dell'Istat.

[3] Il database linkato CeDAP+Campionaria, infatti, ha alcune informazioni che sono presenti in entrambe le fonti originarie: ciò ha consentito di confrontare, controllare e correggere alcune variabili dei CeDAP che presentavano problemi di qualità che impedivano la costruzione delle classi.

[4] Nel 2003 non sono disponibili i CeDAP del Molise, della Calabria e della provincia autonoma di Bolzano. Tra le regioni di cui sono disponibili i CeDAP alcune hanno limiti di copertura della rilevazione, altre di mancanza di informazione, casi mancanti, numero esiguo di casi nelle variabili necessarie alla costruzione delle classi di Robson. Le analisi, pertanto, sono state effettuate solo sulle 12 regioni con qualità del dato migliore: Piemonte, Lombardia, provincia autonoma di Trento, Veneto, Friuli Venezia Giulia, Liguria, Emilia Romagna, Toscana, Umbria, Campania, Puglia, Basilicata per un totale di 23.383 casi

standardizzazione detto "dei coefficienti tipo" che consente di osservare quanti cesarei ci sarebbero in una certa regione se alla sua struttura per classe si applicasse la propensione al cesareo espressa da coefficienti tipo e quanto sarebbe in tal caso il tasso di TC complessivo. Per ogni classe è stato scelto come coefficiente tipo il tasso minimo osservato tra le regioni italiane analizzate, assumendo che i coefficienti così scelti come "tipo" siano dei coefficienti "virtuosi". Ciò anche perché a livello internazionale non esistono raccomandazioni che indichino dei livelli soglia o dei valori ottimali per i tassi di ciascuna classe, data la differenza di contesti e la variabilità della popolazione ostetrica.

3. Risultati

Il ricorso al parto cesareo presenta una forte variabilità territoriale tra le regioni considerate come risulta dalla prima colonna della Tabella 1: è minimo nel Friuli Venezia Giulia (21,4%) ed è massimo in Campania (62,1%).

Tabella 1 – *Tasso di TC e distribuzione della popolazione ostetrica secondo le variabili necessarie alla costruzione delle classi di Robson per regione*

		Parità		Cesareo pregresso		Presentazione del neonato				Età gestazionale		Genere del parto		Modalità del travaglio		
	% TC	Nullipara	Pluripara	No	Almeno uno	Cefalica	Podalica	Obliqua anomal	nd	<=36	>=37	Semplice	Plurimo	Spont.	Indotto	Assente
Piemonte	29,0	53,9	46,1	83,2	16,8	94,6	4,6	0,8		8,4	91,6	98,0	2,0	65,4	15,5	19,2
Lombardia	25,8	56,0	44,0	77,6	22,4	93,1	4,8	0,6	1,5	7,7	92,3	98,0	2,0	65,5	18,9	15,6
P. aut. Trento	27,3	45,1	54,9	77,7	22,3	95,2	4,5	0,3	0,1	8,2	91,8	97,9	2,1	45,6	36,4	18,0
Veneto	29,9	51,0	49,0	84,7	15,3	94,3	4,0	0,3	1,4	9,9	90,1	96,7	3,3	58,6	23,7	17,7
Friuli V. G.	21,4	54,7	45,3	84,0	16,0	96,9	2,9	0,3	0,0	7,2	92,8	98,4	1,6	73,2	14,7	12,1
Liguria	32,9	53,8	46,2	82,3	17,7	93,2	4,7	1,1	1,0	9,3	90,8	98,7	1,3	65,1	14,1	20,8
Emilia Rom.	29,4	56,5	43,5	80,3	19,7	94,5	4,0	0,5	1,1	8,1	91,9	97,8	2,2	66,3	18,7	15,1
Toscana	26,1	56,1	43,9	85,5	14,5	95,0	3,8	0,6	0,7	8,8	91,2	97,1	2,9	67,9	17,0	15,1
Umbria	28,5	57,0	43,0	82,5	17,5	93,4	3,3	0,6	2,7	6,4	93,6	97,9	2,1	71,9	13,4	14,7
Campania	62,1	46,4	53,6	58,0	42,0	90,8	3,7	1,0	4,5	12,0	88,0	97,5	2,6	63,9	1,6	34,6
Puglia	44,4	45,7	54,3	74,7	25,3	93,1	4,4	0,5	1,9	11,4	88,6	96,9	3,1	64,3	15,7	20,1
Basilicata	48,5	46,3	53,8	65,0	35,0	95,4	4,2	0,4	0,1	9,9	90,1	98,6	1,4	56,5	14,4	29,1

Le colonne successive della Tabella 1 riportano la distribuzione della popolazione ostetrica secondo le variabili necessarie alla costruzione delle classi di Robson. Tra le donne che hanno partorito nel 2003 la percentuale di pluripare varia tra il 43% e il 54,9%. Le variabili con campo di variazione più ampio sono: aver già sperimentato un cesareo (14,5%-42%), avere un travaglio spontaneo (45,6%-73,2%) e indotto (1,6%-36,4%). La presentazione cefalica si riscontra in una percentuale di gravidanze che varia tra il 90,8% e il 96,9% e la presentazione podalica in una percentuale di casi fra il 2,9% e il 4,8%. La percentuale di gravidanze terminate entro le 36 settimane, infine, varia dal 6,4% al 12%.

La classificazione di Robson per le regioni prese in esame è presentata nella Tabella 2 che riporta per ogni classe: il peso relativo della classe stessa, il tasso di TC, il contributo assoluto e il contributo relativo al tasso di TC complessivo. Tra parentesi sono indicati il valore minimo e massimo assunti nelle regioni analizzate.

Le classi più numerose (il 30% e il 26% rispettivamente) sono la 1 e la 3 (rispettivamente le nullipare e le pluripare con gravidanza singola terminata ad almeno 37 settimane, presentazione cefalica e travaglio spontaneo) che includono il 56% della popolazione ostetrica analizzata in questo studio. La terza classe in ordine di numerosità è la 2 (nullipare,feto singolo, cefalico, a termine, travaglio indotto o TC prima del travaglio) che comprende il 14% della popolazione ostetrica. Segue poi la 5 (precedente cesareo, feto singolo, cefalico, a termine) che include il 9% dei casi.

E' da sottolineare la forte variabilità regionale della percentuale di TC in alcune classi di Robson: il rapporto fra massimo e minimo è pari a 13,5 volte nella classe 3, a 9,2 volte nella classe 4, a 5,6 volte nella classe 1, a 2,8 nella classe 2.

Le classi che maggiormente contribuiscono all'elevato tasso di TC osservato nel 2003 sono, nell'ordine, la 5, la 2 e la 1, responsabili rispettivamente del 23%, del 21% e del 15% dei cesarei (59% dei parti cesarei complessivamente). Nella classe 5 (precedente cesareo, feto singolo, cefalico, a termine) il tasso di TC è molto elevato (85%) e questo spiega il suo contributo consistente al complesso di cesarei nonostante non sia particolarmente numerosa; nelle regioni italiane considerate i tassi di TC della popolazione ostetrica appartenenti alla classe 5 vanno dal 63% (Friuli Venezia Giulia) al 98% (Campania). La classe 2 (nullipare, feto singolo, cefalico, travaglio indotto o TC prima del travaglio) ha un tasso di TC pari a 50%, con range molto ampio (il tasso varia tra il 34% della Lombardia e il 96% della Campania). Nella classe 1 (nullipare, feto singolo, cefalico, a termine, travaglio spontaneo) il tasso di TC del 17% è relativamente elevato se si tiene conto del fatto che non ci sono condizioni di rischio; anche in questo caso il valore medio è il risultato di forti differenze tra le regioni: il valore minimo (7%) è rilevato in Friuli Venezia Giulia, il valore massimo (40%) in Campania. Invece le donne appartenenti alla classe 3 (pluripare escluso cesareo precedente, feto singolo, cefalico, a termine, travaglio spontaneo), hanno una frequenza di cesarei molto contenuta (6,5%), e pur essendo una delle classi più numerose, contribuisce solo al 5% dei cesarei complessivi. Le classi 6-7-8 (nullipare e pluripare con singolo podalico e gravidanze multiple) hanno dei tassi di TC molto elevati, rispettivamente 96%, 95% e 79% dovuti alle particolari condizioni ostetriche delle donne appartenenti; tuttavia il loro contributo al tasso di TC globale è contenuto (16%) poiché sono classi di dimensione ridotta (comprendono tutte insieme il 6% della popolazione ostetrica). Si noti che i range dei tassi di TC delle classi 6 e 7 sono molto meno ampi che nelle altre classi.

Tabella 2 – *Classi di Robson – analisi regionale*

Classe	Peso relativo della classe (%, range)	Tasso di TC (%, range)	Contributo assoluto al tasso di TC (%, range)	Contributo relativo al tasso di TC (%, range)
1. Nullipare, singolo cefalico, >= 37 settimane, travaglio spontaneo	30,2 (16,1-39,3)	17,4 (7,2-40,5)	5,2 (1,4-11,2)	15,4 (5,2-18,3)
2. Nullipare, singolo cefalico, >= 37 settimane, travaglio indotto o TC prima del travaglio	14,2 (10,8-23,4)	50 (34,2-95,9)	7,1 (5,1-10,3)	20,9 (16,8-31,8)
3. Pluripare (escl. precedente TC), singolo cefalico, >= 37 settimane, travaglio spontaneo	25,6 (21,5-29,3)	6,5 (1,3-17,5)	1,7 (0,3-3,8)	4,9 (1,2-7,4)
4. Pluripare (escl. precedente TC), singolo cefalico, >= 37 settimane, travaglio indotto o TC prima del travaglio	6,9 (4,2-13,6)	37,4 (9,4-86,8)	2,6 (1,3-4,1)	7,5 (4,7-12,0)
5. Precedente cesareo, singolo cefalico, >= 37 settimane	9,2 (5,7-18,4)	84,7 (62,6-97,8)	7,8 (4,0-18,0)	22,8 (16,8-29,4)
6. Tutte le nullipare con singolo podalico	2,4 (1,4-2,9)	96 (88,2-100)	2,3 (1,4-2,8)	6,7 (3,1-10,9)
7. Tutte le pluripare con singolo podalico (incl. precedente TC)	1,4 (0,6-2,3)	95 (89,2-100)	1,3 (0,6-2,2)	3,9 (2,0-6,8)
8. Tutte le gravidanze multiple (incl. precedente TC)	2,5 (1,3-3,3)	79,2 (67,4-100)	2 (1,1-3,0)	5,7 (2,8-8,6)
9. Tutti i singoli con presentazione anomala (incl. precedente TC)	0,6 (0,2-1,0)	82,9 (55,3-100)	0,5 (0,2-1,0)	1,4 (0,5-2,2)
10. Tutti i singoli cefalici, <= 36 settimane (incl. precedente TC)	7,2 (4,8-10,2)	50,9 (32,8-74,4)	3,7 (1,7-7,6)	10,8 (6,1-12,4)
Totale	100,0	34,0 (21,4-61,3)	34,0	100,0

L'analisi è proseguita standardizzando i tassi di TC delle 12 regioni analizzate. La prima colonna della Tabella 3 riporta la serie teorica dei tassi virtuosi di TC utilizzata come coefficienti tipo: le regioni che presentano i "tassi più virtuosi" sono quelle del Nord Est, Umbria e Toscana.

Si è scelto di presentare e commentare i risultati della standardizzazione relativi alle due regioni con tasso di TC più basso e più elevato, ossia Friuli Venezia Giulia e Campania. Tali risultati sono mostrati nella Tabella 3 che riporta per ogni classe il peso relativo della classe, il tasso di TC reale e il rapporto tra numero di cesarei teorico e numero di cesarei realmente osservati; per ognuna delle due regioni, inoltre, la tabella riporta il tasso complessivo di TC reale, quello teorico (ossia standardizzato) e il rapporto tra tasso teorico e tasso reale. I rapporti tra numero di cesarei teorico e reale nelle diverse classi forniscono un'indicazione sulla riduzione del numero di cesarei che si avrebbe in ogni classe se alla popolazione ostetrica della regione venisse applicata una propensione "virtuosa" a ricorrere al parto cesareo.

Applicando i "tassi virtuosi" nelle due regioni si avrebbe un decremento del numero, e quindi del tasso di cesarei, lieve (20%) in Friuli Venezia Giulia, forte (56%) in Campania. In entrambe le regioni si osserva che il decremento è contenuto quando il feto di gravidanza singola si presenta in posizione podalica (classi 6 e 7) o si tratta di gravidanza multipla (classe 8), mentre è più forte quando si applicano i "tassi virtuosi" in qualcuna delle prime quattro classi: in Campania il numero di cesarei si ridurrebbe ad un decimo nelle classi 1-3 e 4 (i gruppi che non presentano particolari fattori di rischio), seguite dalla 2 e dalla 10 (i pretermine). E' comunque interessante notare che in Campania, nonostante l'enorme riduzione, il tasso teorico di TC resta comunque più elevato di quanto si riscontra in Friuli

Venezia Giulia. Questa differenza è attribuibile alla classe 5 (almeno un cesareo precedente) che include oltre il 18% della popolazione di interesse in Campania, contro il 6% in Friuli Venezia Giulia.

Tabella 4 – *Risultati della standardizzazione*

Classe	Tasso di TC teorico		FRIULI VG			CAMPANIA		
			Peso relativo	Tasso di TC reale	N. Cesarei teorico/reale	Peso relativo	Tasso di TC reale	N. Cesarei teorico/reale
1.	7,2	Friuli VG	36,3	7,2	**1,00**	27,6	40,5	**0,18**
2.	34,2	Lombardia	12,1	47,3	**0,72**	10,8	95,9	**0,36**
3.	1,3	P. aut. Trento	29,3	2,0	**0,66**	21,5	17,5	**0,08**
4.	9,4	P. aut. Trento	5,3	35,3	**0,27**	4,2	86,8	**0,11**
5.	62,6	Friuli VG	6,4	62,6	**1,00**	18,4	97,7	**0,64**
6.	88,2	Emilia R	1,4	100,0	**0,88**	1,9	100,0	**0,88**
7.	89,2	Umbria	1,2	100,0	**0,89**	1,6	93,3	**0,96**
8.	67,4	Lombardia	1,6	82,3	**0,82**	2,7	88,9	**0,76**
9.	55,3	Veneto	0,2	86,0	**0,64**	1	100,0	**0,55**
10.	32,8	Toscana	6,2	40,2	**0,82**	10,2	74,4	**0,44**
Tasso di TC reale				21,4			61,3	
Tasso di TC teorico				17,2			26,7	
Tasso di TC teorico/reale				**0,80**			**0,44**	

4. Discussione e conclusioni

Anche nelle regioni italiane dove il ricorso al TC è meno frequente, i tassi superano abbondantemente le soglie massime del 10-15% raccomandata dal WHO, e del 15-20% indicata dal Ministero della Salute.

Le classi di Robson maggiormente responsabili sia dell'alto tasso di TC complessivo che delle differenze regionali di TC sono: a) quelle teoricamente a minor rischio, che includono in tutte le regioni una percentuale molto elevata delle nascite, ma per le quali ci sono grandi differenze di tasso di TC; b) il cesareo pregresso, per il quale è molto forte la variabilità regionale per quanto riguarda sia la frequenza nella popolazione che il tasso di TC. Questo risultato è comune ad altri studi relativi a contesti in cui il ricorso al TC è aumentato molto velocemente negli ultimi anni (Betran et al, 2014).

Mentre per le classi a basso rischio è evidente il ruolo delle pratiche ostetriche nella determinazione della variabilità regionale del tasso di TC, per quanto riguarda il cesareo pregresso hanno un ruolo importante anche le caratteristiche della popolazione. I livelli di fecondità del passato (complessiva e per ordine di nascita) determinano sia la proporzione di nullipare/pluripare che, in parte, la frequenza di donne che hanno già sperimentato un TC: nelle regioni del sud, con fecondità più elevata, c'è una maggior percentuale di pluripare con più di due figli, e quindi è maggiore la parte di popolazione a rischio di aver già avuto almeno un cesareo.

Tuttavia, sia le forti differenze riscontrate nella percentuale di popolazione ostetrica con cesareo pregresso, che il livello del tasso di TC, dipendono a loro volta dalle pratiche medico sanitarie: in una regione medicalizzata come la Campania, dove il ricorso al TC è elevatissimo e radicato nella pratica medica ormai da anni, le donne con figli molto probabilmente hanno già partorito almeno una volta con cesareo e lo ripeteranno ancora, alimentando la diffusione del fenomeno. Si tenga presente che anche se le indicazioni sulle modalità di travaglio e di parto più appropriate in caso di pregresso cesareo sono ancora oggi oggetto di dibattito, alcune società scientifiche raccomandano la proposta di un travaglio di prova nei casi di pregresso cesareo senza specifiche controindicazioni, previa discussione con la donna di rischi e benefici che questa opzione comporta (Dodd et al, 2004).

A conferma del ruolo prevalente delle pratiche ostetriche si ricorda che la variabilità del tasso di TC tra le regioni è molto ampia sia nelle classi senza particolari condizioni di rischio, che nelle gravidanze multiple, in quelle pretermine e per le donne con cesareo pregresso. Questo conferma la disomogeneità e la non appropriatezza delle pratiche cliniche e assistenziali adottate che era stata ipotizzata.

Per limitare la pratica del TC in Italia due interventi si raccomandano particolarmente: ridurre drasticamente la pratica del TC per le donne al primo figlio senza fattori di rischio, proporre il travaglio di prova alle donne con pregresso cesareo che non abbiano specifiche controindicazioni.

Riferimenti bibliografici

BETRAN AP, MERIALDI M, et al, 2007. Rates of caesarean section: analysis of global, regional and national estimates, *Paediatr Perinat Epidemiol*, Vol. 21, No.2, pp.98-113.

BETRÁN AP, VINDEVOGHEL N, et al, 2014. A Systematic Review of the Robson Classification for Caesarean Section: What Works, Doesn't Work and How to Improve It, PLoS ONE, Vol. 9, No. 6.

DODD JM, CROWTHER CA, et al, 2004. Planned elective repeat caesarean section versus planned vaginal birth for women with a previous caesarean birth. *The Cochrane Database of Systematic Reviews*, Issues 4.

ISTAT, 1980. *Annuario di statistiche demografiche*. Roma.

ISTITUTO SUPERIORE DI SANITÀ 2012, Taglio cesareo una scelta appropriata e consapevole, Linee guida n.22, Roma.

LAVENDER T, HOFMERY GJ, et al, 2006. Caesarean section for non medical reasons at term, *Cochrane Database of Systematic Reviews*, Issue 3.

MINISTERO DELLA SALUTE, 2013. *Rapporto annuale sull'attività di ricovero ospedaliero Dati SDO 2012*. Roma.

MINISTERO DELLA SALUTE, 2013a. *Certificati di Assistenza al Parto Anno 2010*. Roma.

REGIONE EMILIA ROMAGNA, 2013. *La nascita In Emilia Romagna - 10° Rapporto sui dati del Certificato di Assistenza al Parto (CeDAP) - Anno 2012.* Bologna.

ROBSON MS, 2001. Can we reduce the caesarean section rate?, *Best Pract Res Clin Obstet Gynaecol*, Vol.15, No. 1, pp. 179-194.

VILLAR J, VALLADARES E, et al, 2006. Caesarean delivery rates and pregnancy outcomes: the 2005 WHO global survey on maternal and perinatal health in Latin America, *The Lancet*, Vol. 367, No. 9525, pp. 1819-1829.

WHO, 1985. Appropriate technology for Birth, *The Lancet*, Vol. 2, pp.436-437.

SUMMARY

Appropriateness of Caesarean Section in the Italian regions: analysis of Caesarean Section rate by using Robson's classification

The aim of this paper is to compare the variability and appropriateness of the use of Cesarean Section (CS) in Italian regions by means of the "Robson's classification", which provides a standard framework for monitoring, auditing and analyzing the cesarean section rate: in particular to identify which groups of women are contributing most to the high CS rate in the different regions; and to analyze whether the variability of the overall CS rate between regions is dependent on the different structure of the obstetric population, rather than on differences in obstetric practice. The results show that in order to limit the practice of cesarean section in Italy, public health interventions are recommended, so as to reduce the cesarean section among nulliparous women without risk factors and to promote trial of labor for women with a previous cesarean section.

Francesca LARICCIA, Istituto Nazionale di Statistica, lariccia@istat.it
Antonella PINNELLI, Sapienza Università di Roma,pinnelli.antonella@gmail.com
Sabrina PRATI, Istituto Nazionale di Statistica, prati@istat.it
Marina ATTILI, Istituto Nazionale di Statistica, maattili@istat.it
Claudia IACCARINO, Istituto Nazionale di Statistica, iaccarin@istat.it

Rivista Italiana di Economia Demografia e Statistica Volume LXVIII n. 3/4 Luglio-Dicembre 2014

UN'ANALISI GEOGRAFICA SULLA PRESENZA DEI CITTADINI STRANIERI A ROMA

Federico Benassi, Fabio Lipizzi, Donatella Zindato

1. Contesto di riferimento, obiettivi e struttura del lavoro

L'aumento della popolazione nel decennio intercensuario 2001-2011 (+2,4 milioni) è frutto, com'è noto, di un notevole aumento della componente straniera (+2,7 milioni), a fronte di una lieve contrazione di quella italiana (-255 mila). L'intensità della crescita della componente più stabile della popolazione straniera ben si comprende se pensiamo che al 2001 il suo peso relativo sul totale della popolazione residente superava di poco i 2 punti percentuali (2,3%), mentre al 2011 sfiora i 7 (6,8%). Un aspetto ad oggi meno dibattuto è la dimensione spaziale di tale aumento che, comprensibilmente, ha interessato il territorio italiano in modo disomogeneo. All'interno di tale contesto, assume particolare rilevanza l'impatto della popolazione straniera sulle dinamiche demografiche dei grandi comuni (Benassi *et al.*, 2014; Strozza *et al.*, in corso di stampa). Il caso di Roma è significativo per molteplici ragioni: città primate della struttura urbana italiana, è caratterizzata da una spiccata eterogeneità interna, in termini morfologici e territoriali, nonché dalla presenza di una molteplicità di collettività straniere sia di antico che di più recente insediamento (Conti e Strozza, 2006; Crisci, 2010; Cristaldi, 2002). Inoltre, è uno dei grandi comuni per i quali si registra un'incidenza (stranieri per cento residenti) maggiore di quella media nazionale, sia al 2001 che al 2011 (rispettivamente, 3,9% contro 2,3% nel 2001 e 8,6% contro 6,8% nel 2011). Il contributo si propone un duplice obiettivo: presentare una prima serie di applicazioni di analisi geografica sulla presenza straniera nel comune di Roma ai censimenti 2001 e 2011; mostrare le potenzialità dei dati censuari per lo studio dei fenomeni socio-demografici con approccio geografico. La struttura del lavoro è la seguente: nel paragrafo 2 sono descritti i dati, il dettaglio territoriale di analisi e i metodi utilizzati; nel paragrafo 3 viene analizzata la distribuzione micro-territoriale dei cittadini stranieri ai censimenti 2001 e 2011; nel paragrafo 4 vengono presentate alcune riflessioni conclusive.

2. Dati, dettaglio territoriale e metodi di analisi

I dati utilizzati si riferiscono all'insieme della popolazione residente nel comune di Roma ai censimenti 2001 e 2011, distinta per cittadinanza italiana o straniera. Il collettivo dei cittadini stranieri è stato considerato come un insieme omogeneo, in quanto i dati relativi alla distribuzione per sezione di censimento e singola cittadinanza al 2011 non sono ancora disponibili. Ciò rappresenta una limitazione rilevante alla portata informativa del contributo realizzato; è noto, infatti, che le diverse collettività straniere tendono ad insediarsi in modo differenziato sul territorio dando vita ad una gamma relativamente ampia di modelli insediativi, così come è stato evidenziato da alcuni studi condotti sul caso italiano (Benassi e Ferrara, 2013; Ferrara *et al.*, 2010; Ferruzza *et al.*, 2008). Il dettaglio territoriale di analisi sono dunque le sezioni di censimento e le loro aggregazioni in aree sub-comunali. La metodologia adottata rientra nell'ambito dell'analisi spaziale e, in particolare, delle statistiche centrografiche. Partendo dalla distribuzione per sezione di censimento degli stranieri residenti nel comune di Roma al 2001 e al 2011, sono stati calcolati il centro medio semplice, il centro medio pesato e la deviazione standard ellittica. Il centro medio semplice è un punto fittizio identificato su un piano euclideo attraverso il calcolo delle medie semplici delle coordinate geografiche (longitudine e latitudine) dei punti di un determinato territorio (nel nostro caso, i centroidi di ciascuna sezione di censimento del comune di Roma). E' dunque un indice sintetico, ottenuto sotto l'ipotesi di assenza di variabilità nella distribuzione della popolazione, e rappresenta il punto di equilibrio teorico che si realizzerebbe qualora sussistesse una condizione di perfetta equidistribuzione della popolazione (nel nostro caso straniera) tra le sezioni di censimento del comune di Roma. Indicando con X la longitudine e con Y la latitudine, il centro medio semplice di un qualsivoglia territorio composto da *n* poligoni (nel nostro caso, da *n* sezioni di censimento) avrà come longitudine (Xs) la media aritmetica delle longitudini degli *n* centroidi e come latitudine (Ys) la media aritmetica delle latitudini. In formula:

$$X_s = \frac{\sum_{i=1}^{n} x_i}{n} \qquad Y_s = \frac{\sum_{i=1}^{n} y_i}{n} \qquad (1)$$

Il centro medio pesato è anch'esso un indice sintetico che, ottenuto come media delle coordinate dei singoli centroidi ponderate per la loro numerosità (il loro peso demografico), tiene conto delle diverse dimensioni demografiche dei punti (ovvero, nel nostro caso, della numerosità della popolazione straniera delle sezioni di censimento). Mantenendo la notazione proposta per la [1] e indicando il peso

demografico con *p*, il centro medio pesato avrà dunque le seguenti coordinate geografiche:

$$X_p = \frac{\sum_{i=1}^{n} x_i p_i}{\sum_{i=1}^{n} p_i} \qquad Y_p = \frac{\sum_{i=1}^{n} y_i p_i}{\sum_{i=1}^{n} p_i} \tag{2}$$

La misura ottenuta in questo secondo caso è dunque sensibile alle variazioni di peso dei punti. Per questo motivo la ricostruzione della sua evoluzione temporale permette di individuare le direttrici dei processi di ridistribuzione territoriale sperimentati da una data popolazione in un determinato arco temporale (nel nostro caso il decennio 2001-2011). Infine, la deviazione standard ellittica è una misura che qualifica la portata informativa del centro medio semplice e pesato, in quanto è in grado di misurare la dispersione intorno al centro medio da un punto di vista geografico. Infatti, scomponendo la distanza quadratica dal centro medio (semplice o pesato), si ottengono due componenti, σ_x e σ_y, che indicano la dispersione lungo gli assi est-ovest e nord-sud della variabile osservata (David e Wong, 2001). Tra le infinite coppie di assi ortogonali passanti per il centro medio, vengono scelte quelle che hanno una correlazione nulla tra le coordinate e che massimizzano uno dei due scarti quadratici medi, minimizzando il secondo. Gli assi si ottengono ruotando il sistema di un angolo α dalla (3):

$$tg\,2\alpha = \frac{2Cov(x_i, y_i)}{\alpha_x^2 - \alpha_y^2} \tag{3}$$

dove, per $\sigma_x \neq \sigma_y$, la covarianza tra le coordinate x_i e y_i è:

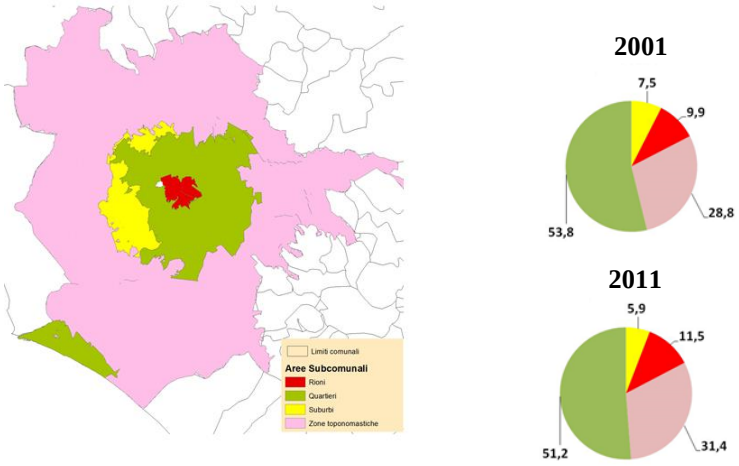
$$Cov(x_i, y_i) = \frac{\sum_{i=1}^{N} \left(x_i - X_p\right)\left(y_i - Y_p\right)}{\alpha_x^2 - \alpha_y^2} \tag{4}$$

Si può allora generare un'ellisse con centro nelle coordinate del centro medio semplice o pesato, dove l'asse maggiore è il valore più alto tra σx e σy e l'asse minore è il valore più piccolo tra σx e σy. Gli assi dell'ellisse consentono di interpretare la variabilità geografica dei punti analizzati. Infatti, tanto più la forma si presenta allungata tanto maggiore è la dispersione sul territorio del fenomeno in esame, sia rispetto alla latitudine (asse nord-sud) che alla longitudine (asse est-ovest). L'angolo di rotazione, inoltre, indica la direzione geografica privilegiata lungo la quale il fenomeno si manifesta (Cruciani *et al*., 2011).

3. La distribuzione micro territoriale dei cittadini stranieri a Roma

Un primo aspetto affrontato è stato quello della distribuzione percentuale dei cittadini stranieri nelle aree sub-comunali in cui si suddivide il territorio capitolino. La suddivisione storico-toponomastica, ricostruibile anch'essa a partire dalle sezioni di censimento (Lipizzi, 2013), ha l'indubbio vantaggio di consentire il riferimento al territorio attraverso una denominazione perfettamente riconoscibile; infatti, è possibile identificare la localizzazione geografica di un'area sia attraverso la sua tipologia (rioni, quartieri, suburbi e zone toponomastiche - Figura 1) sia con il suo toponimo (ad. es. Rione Monti o Trastevere). Tra il 2001 e il 2011, la quota di stranieri censiti sale sia nei rioni, da circa il 10% a 11,5%, che nelle zone toponomastiche, da poco meno del 29% ad oltre il 31%. Suburbi e quartieri vedono al contrario contrarre la quota di stranieri ivi residenti, che scende rispettivamente da 7,5% a circa il 6% e da 53,8% a 51,2% (Figura 1).

Figura 1 – *Le aree sub comunali di Roma. Distribuzione percentuale degli stranieri residenti nelle aree sub comunali di Roma. Censimenti 2001-2011.*



Quartieri = 34; Rioni = 22; Suburbi = 6;
Zone toponomastiche = 53

Fonte: nostre elaborazioni su dati Istat

Sembra quindi che nel decennio 2001-2011 sia aumentata la quota di stranieri residenti nel centro della città ma, soprattutto, la quota di stranieri residenti nelle aree relativamente meno centrali del territorio comunale. Per chiarire la natura della presenza straniera nel centro storico sarà necessario attendere la disponibilità dei dati per singola cittadinanza, mentre l'aumento di stranieri residenti nelle aree

più lontane dal centro sembra indicare una maggiore dispersione sul territorio della popolazione straniera. Ulteriori elementi che sembrano confermare la tendenza alla dispersione della presenza straniera sul territorio comunale emergono dall'analisi delle Figure 2 e 3 e dalla Tabella 1.

Figura 2 – *Distribuzione territoriale degli stranieri residenti per sezione di censimento. Censimenti 2001-2011.*

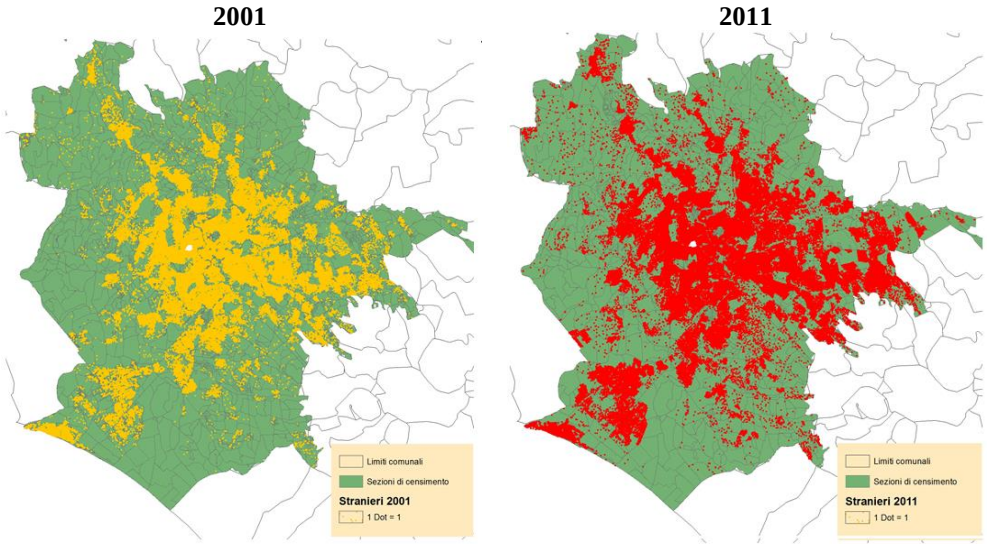


Tabella 1 – *Alcuni indicatori sulla distribuzione territoriale degli stranieri residenti per sezioni di censimento. Censimenti 2001-2011*

Indicatori	**2001**	**2011**
Numero stranieri censiti	98mila	224mila
Sezioni di censimento con almeno un cittadino straniero censito	8.912	10.433
Sezioni di censimento con almeno uno straniero censito per 100 sez. di censimento	65,9%	77,2%

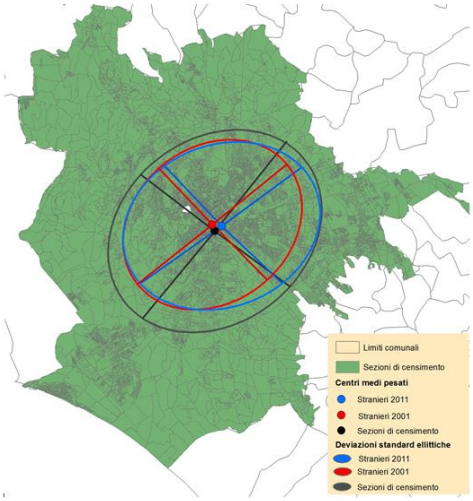
Fonte: nostre elaborazioni su dati Istat

Nella Figura 2 è rappresentata la distribuzione micro-territoriale - a livello di sezione di censimento - dei cittadini stranieri al 2001 e al 2011; in entrambi i cartogrammi, a un cittadino straniero corrisponde un punto sulla mappa. Da un'analisi visiva, appare abbastanza evidente una certa espansione territoriale dei punti che, nel 2011, vanno a "riempire" porzioni di territorio non interessate (o solo

parzialmente interessate) dalla presenza straniera al 2001. Ciò è confermato dal parallelo aumento, oltre che del numero di stranieri residenti (A), del numero di sezioni con almeno un cittadino straniero (B) e del loro peso relativo (C) (Tabella 1).

La Figura 3 chiarisce definitivamente e qualifica da un punto di vista geografico quanto detto fin qui. Se è vero infatti che la distribuzione geografica dell'insieme dei cittadini stranieri residenti a Roma è diversa da quella attesa nell'ipotesi di equa distribuzione spaziale (ellisse nera), sia nel 2001 (ellisse rossa) che nel 2011 (ellisse blu), è altrettanto vero che il cambiamento nella forma della ellissi di colore blu indica proprio una maggiore diffusione sul territorio comunale della componente straniera; maggiore diffusione che, osservando il verso della rotazione ellittica, interessa in particolare il quadrante orientale del territorio comunale.

Figura 3 – *Centro medio, centro medio pesato e deviazioni standard ellittiche per gli stranieri residenti. Censimenti 2001-2011*



Fonte: nostre elaborazioni su dati Istat

4. Conclusioni e sviluppi futuri

L'analisi condotta ha mostrato le potenzialità dei dati censuari per lo studio e l'analisi dei fenomeni demosociali da un punto di vista micro-territoriale con approccio geografico e costituisce un primo tentativo di indagare le recenti dinamiche territoriali della presenza straniera a Roma. In particolare, si è

evidenziato come al 2011 gli stranieri risultino meno concentrati rispetto al 2001 (anche se permane una situazione distante da quella teorica di equa distribuzione areale). Tale diffusione territoriale sembra aver interessato soprattutto il quadrante orientale del territorio capitolino. Si tratta di un'evidenza coerente con note impostazioni teoriche secondo cui, con il passare del tempo, la presenza straniera tenderebbe a ridurre progressivamente il proprio grado di concentrazione territoriale, uniformandosi alla distribuzione degli autoctoni, o comunque a disperdersi sul territorio. Naturalmente, al fine di acquisire una maggiore comprensione delle dinamiche territoriali che hanno interessato la presenza straniera a Roma, è necessario approfondire l'analisi e qualificare questi primi risultati attraverso l'utilizzo dei dati relativi alla distribuzione micro-territoriale delle singole collettività.

Riferimenti bibliografici

BENASSI F., FERRARA R. 2013. Modelli insediativi delle principali collettività immigrate in Italia: recenti tendenze, *Rivista di Economia e Statistica del Territorio,* 2, pp. 66-85.

BENASSI F., FERRARA R., GALLO G., STROZZA S. 2014. Il contributo degli immigrati alla dinamica demografica di alcuni ambiti urbani italiani, *Rivista Italiana di Economia, Demografia e Statistica*, LXVII (1), pp. 47-54.

CONTI C., STROZZA S. 2006. *Gli immigrati stranieri e la capitale*, Franco Angeli, Milano.

CRISTALDI F. 2002. Multiethnic Rome: Toward residential segregation?, *GeoJournal*, 58, pp.81-90.

CRUCIANI S., LIPIZZI F., MUGNOLI S., ARCASENZA M., ENDENNANI G. 2011. Una stima 2001-2011 dell'"urban sprawl" in Italia attraverso l'uso di dati geografici in AISRe (a cura di) XXXII Conferenza Italiana di Scienze Regionali. Istituzioni, Reti Territoriali e Sistema Paese: la governance delle relazioni locali.

CRISCI M. 2010. *Italiani e stranieri nello spazio urbano. Dinamiche della popolazione di Roma*, Franco Angeli, Milano.

DAVID J.L., WONG W.S. 2001. *Statistical Analysis with ArcView Gis*, John Wiley & Sons, New York.

FERRARA R., FORCELLATI R., SROZZA S. 2010. Modelli insediativi degli immigrati stranieri in Italia, *Bollettino della Società Geografica Italiana*, 13(3), pp. 619-639.

FERRUZZA A., DARDANELLI S., HEINS F., VERRASCINA M. 2008. La geografia insediativa degli stranieri residenti: Verona, Firenze e Palermo a confronto, *Studi Emigrazione*, 171, pp. 601-608.

LIPIZZI F. 2013. *Innovazioni di processo e di prodotto nelle fasi di aggiornamento delle basi territoriali 2010-2011*. Istat Working Papers n 2/2013. Istat– Roma.
STROZZA S., BENASSI F., FERRARA R., GALLO G. La recente evoluzione demografica nei maggiori ambiti urbani italiani e il fondamentale ruolo degli stranieri, in corso di stampa su *Archivio di Studi Urbani e Regionali*.

SUMMARY

A geographical analysis of foreign presence in Rome[1]

The paper focuses on the use of centrographic techniques for the spatial analysis of demographic phenomena. The territorial dynamics of foreign presence in Rome in-between the past two censuses have been explored by comparing the territorial distribution of foreign citizens at the enumeration area level. The first evidence coming from the analysis, i.e. a general tendency to dispersion, need to be further investigated when data by country of citizenship will be released.

Federico BENASSI, Istituto Nazionale di Statistica, benassi@istat.it
Fabio LIPIZZI, Istituto Nazionale di Statistica, lipizzi@istat.it
Donatella ZINDATO, Istituto Nazionale di Statistica, zindato@istat.it

Rivista Italiana di Economia Demografia e Statistica Volume LXVIII n. 3/4 Luglio-Dicembre 2014

TRIGGER FACTORS THAT INFLUENCE BANKRUPTCY: A COMPARATIVE AND EXPLORATORY STUDY

Leonardo Di Marco, Luciano Nieddu

1. Introduction

The phenomenon of bankruptcy has significantly prejudiced the Italian productive fabric in the last few years. The impact of the world economic crisis on the Italian economy has generated devastating effects for the wealth of the Nation, causing strong unbalances in terms of employment, productivity and investments.

Business failure prediction is one of the most essential problems in the field of finance. The research on developing business failure prediction models has been focused on building classification models to distinguish among failed and non-failed firms. The pioneer for corporate failure prediction models was William Beaver (1966). He applied a univariate model in which a classification model was carried out separately for each financial ratio, and an optimal cut-off point was identified so that the percentage of misclassifications was minimized.

Altman (1968) applied Linear Discriminant Analysis introduced by Fisher in 1936 to the problem of predicting bankruptcy. This technique dominated the literature on corporate failure models until the 1980s and is commonly used as benchmark for comparative studies.

Altman's study involved 66 manufacturing companies with equal number of failures and survivors, and a total of 22 ratios from five categories, namely liquidity, profitability, leverage, solvency, and activity. From this set of ratios, five were finally chosen on the basis of their predictive ability.

Since these seminal studies not much work has been done to apply different methodologies to the problem of failure predictions, until the early 80s when Ohlson (1980) and Zmijewksi (1984) applied logistic regression to the problem of predicting bankruptcy.

The goal of this paper is twofold: determine if the information obtainable from the financial statement can be used to predict the failure of a company and analyze financial statement items and ratios to find out which variables are the most determinant for the failure or the survival of a company.

This research focuses on the situation of firms located in the province of Lazio and covers a time frame of eleven years, from 2000 to 2011. The study has been carried out on a stratified sample of 100 firms from various economic sectors.

This paper differs from previous works since we used all the items of a financial statement as covariates and not only some performance ratios. Since the data at hand suffers from some serious issues we have decided to apply a very robust non-parametric classification technique such as classification trees (Breiman et al, 1984) in order to select those items that should be helpful in determining the failure of a company.

The paper unfolds as follows: in section 2 a brief description of the main issues related to the available data will be carried out. In section 3 the methodology used to analyze the data will be summarized and the results will be presented in section 4. In section 5 some conclusions will be drawn.

2. Sample's Issues

The stratified sample is composed of 100 companies randomly selected from those that, at year 2000, had revenue from sales between 2 million euros and 50 million euros. Of these firms, 50 were still active at 2011 while 50 had filed for bankruptcy sometime during the considered time frame.

The sample was stratified with respect to economic sector. No firms operating in the financial sector were included in the analysis, since they are known to react and operate differently from the other firms in case of distress. For each company the financial statement was available.

A financial statement describes the activities and the performances of a business throughout a specific period of time. It is composed by three main parts:

a) **Balance sheet**, that provides detailed information about assets, liabilities and shareholders
b) **Income statement** (or profit and loss account) shows the company's revenues and expenses during a specific period of time and it is made of revenues and expenses
c) **Notes to financial statement** are additional information that further detail specific items as well as provide a more comprehensive assessment of a company's financial condition.

In addition to the data of the financial statements, further information can be obtained from the financial statement to evaluate the profitability, solvency, liquidity and stability of a business.

Considering this very short summary of the content of any financial statement, it is clear that, using all the items of a financial statement as variables, yields a

dataset with an enormous number of variables and with a sparse structure in term of non-missing values.

Some other limitations of the data are:

a) Financial statements, prepared by the companies themselves, are the main source of external information, and most of the firm's performance evaluation is based on it. The final financial statement, therefore, may be the result of an adjustment that is performed within the boundaries of existing legislation to make it suitable to the particular and contingent needs of a company. Therefore some balance sheets not only reflect the financial and economic status of the firm but are also the outcome of a particular need that the firm is facing.
b) Although the firms that we analyzed are only small-medium firms, this category includes companies that can present a wide range of values for each item of the balance sheet: some with few employees and low absolute turnover, and some with several employees and high-volume business. This makes the sample data very heterogeneous and diverse.

Considering all these limitations and the small sample size, the application of a robust non-parametric methodology was necessary. Among the many models available in data mining we have opted for classification trees (Breiman et al., 1984), which are known to perform quite well in the presence of missing data and as a tool for feature selection when a large number of variables are available.

3. C&RTs

Classification and Regression Trees are a non-parametric statistical method, conceptually simple and yet powerful. They split the feature space into rectangles and then fit a simple model in each subset. Regression trees deal with continuous outcomes while classification trees consider a polytomous response variable (class) that is supposed to depend on a set of covariates.

They are particularly useful in data mining when there is a plethora of covariates with missing data. They are invariant to monotone transformations of the variables and are very robust with respect to outliers.

The goal of classification trees is to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data at hand that has been previously classified (supervised classification).

Tree-based classification methods split the covariate space X into disjoint set of rectangular regions, and then classify the observations according to the mode of the class of the elements that belong to that region. The partitioning is usually performed according to an impurity measure (usually the Gini index) or according

to the information gain (entropy) that can be achieved once the covariate space has been partitioned. Therefore, starting with a single node (root) containing all the elements in the training data, we look for the binary partition that yields the best information gain or impurity reduction. The data are then partitioned according to the values of that covariate, yielding two subsets that stem from the root and which should be more homogenous according to the outcome than the set that generated them. This partitioning is recursively performed on the derived subsets and it stops either when the units in a node have all the same value of the response variable or when splitting no longer adds value to the predictions or the number of elements in the derived subset decreases under a pre-specified threshold.

The iterative partitioning process is called "growing a tree" or "learning". When dealing with more than one covariate, the one leading to the split with the lowest impurity is first selected. The terminal nodes of the tree are called the leaves. Each leaf is assigned a class according to a majority rule based on the classes of the elements that belong to that leaf. This majority rule criterion is also used in classifying new objects.

Two main issues are connected to classification trees:

- the problem of learning an optimal decision tree is known to be NP-complete therefore decision-tree learning algorithms are based on heuristics such as the greedy algorithm, where locally-optimal decisions are made at each node. Such heuristics cannot guarantee that the results be the globally-optimal decision tree;
- classification trees algorithms can create over-complex trees. The complexity of the tree doesn't necessarily imply a good accuracy of the tree. A too complex tree will clearly perform well on the training data (overfitting), but this not necessarily means that it will be able to correctly classify new objects of unknown class. To avoid over complex trees, pruning techniques usually based on cross validation (i.e. on their performance on new data) can be used.

Classification trees, to our knowledge, have only been applied once to business failure in a study that did not produce reliable results due to a very small sample size (Huarng et al., 2005).

4. Experimental Setup and Results

The collected data refers to balance sheets of companies from 2000 to 2011. For the companies that are still active at 2011 all the balance sheets are available. The study that has been carried out is a cross-sectional study: companies have been considered at various years prior to failure and for each failed company the balance sheet up to eight years prior to failure (if available) has been considered. Each failed company has been matched to a non-failed company that, in the same year,

presented a balance sheet, operated in the same economic sector and was comparable in size. Therefore eight datasets were generated with equal number of failed and non-failed companies.

The performance of each classification tree was assessed via 10-fold cross validation. In Table 1 the resubstitution error rate (R-ER) and cross validation error rate (CV-ER) have been reported.

Table 1 – *resubstitution error rates and cross validation error rates.*
R-ER: apparent error rate
CV-ER: 10-fold crossvalidation error rate.

	YEARS PRIOR TO FAILURE							
	1	***2***	***3***	***4***	***5***	***6***	***7***	***8***
R-ER	0.051	0.051	0.071	0.083	0.078	0.081	0.065	0.029
CV-ER	0.092	0.143	0.102	0.135	0.100	0.081	0.109	0.088

In Figures 1 and 2 some of the eight classification trees have been displayed. One variable that is always influential in determining the failure of a firm for all the trees that have been grown is the "financial charges on sales": companies that are still active at 2011 present a ratio between financial expenses and revenues greater than those that have failed. This could at first seem counterintuitive since financial charges are considered a negative “asset” for the firms. This result must be embedded in the Italian economic framework, where access to the credit system is not as flexible as it should be in a healthy efficient economic system. “Financial charges on sale” can then be considered a proxy of the ability of a firm to access the credit system: high values for this variable indicate the ability of a company to access the credit system and therefore survive even after turnover's reductions.

In Figure 1, the decision tree shows that companies with a percentage of financial charges on sales over 15% are those who remained active, while those with less than 15% failed after eight years. These results are very similar to the seventh, sixth and fifth year preceding the bankruptcy, where the only discriminating variable was the relationship between financial expenses and revenues. The situation becomes more interesting from the fourth year up to one year before the bankruptcy, as other discriminating variables come into play.

Figure 2 shows the tree for one year prior to failure. Only a few years prior to failure the standard performance ratios come into play. Among the companies with the item "financial charges on sales" higher than 7%, 40 are still in business whereas 6 have started bankruptcy procedures: the next subdivision is given by ROS index that measures the return on sales. Firms with a ROS higher than 14,5% remained in activity, but those with a ROS less than 14,5% failed. This is a natural conclusion because the last balance sheet of a company before the bankruptcy

represents a situation where the actual activity of the company is already finished, so the "sales revenue" of these companies has a value of zero.

Figure 1 – *Decision tree 8 years before bankruptcy*

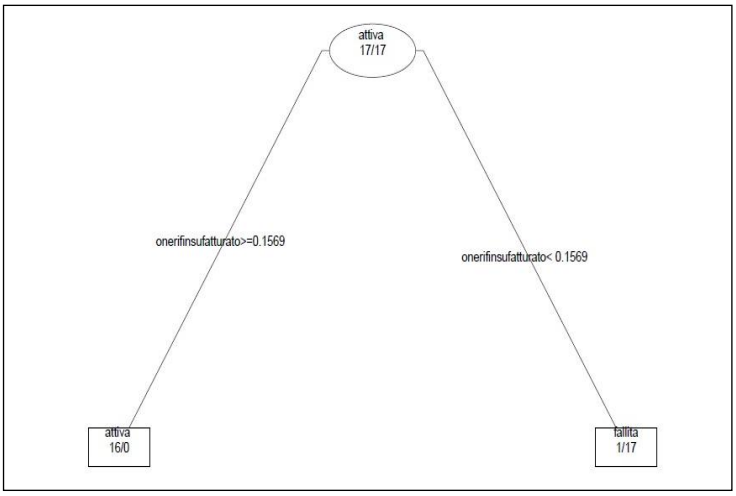
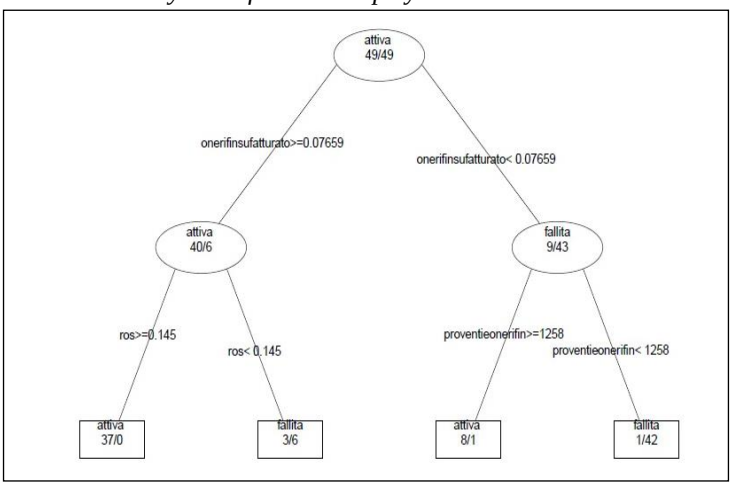


Figure 2 – *Decision tree 1 year before bankruptcy*



On the other hand, companies with a "financial charges on sales" less than 7%, are mostly companies that then have started a bankruptcy procedure (43), whereas few companies remained in business (9). The additional discriminating variable is given by the ‘‘financial proceeds and charges’’: a very positive difference between

financial proceeds and charges has brought the survival of companies, whereas those who have had a minor discrepancy between financial proceeds and charges are going to fail.

The bottom-line of all the classification trees is that some companies continued to have access to credit and this allowed them to stay in business despite the crisis period; on the other hand, other companies were not able to be granted credit by banks and then carry out their production. Hence, they have been compelled, due to a sharp drop in sales resulting from a drastic reduction in production, to declare the state of crisis, and then bankruptcy.

5. Conclusions

In this paper we have used classification trees to predict firm bankruptcy based on all the items of the financial statement and some performance ratios at various years prior to failure. This is an unusual approach to the problem that has not been properly studied in business failure literature. We have found that the discriminant of business failure in Italy has been the inability of the companies to access the credit system in order to continue production. One conclusion is that the timely financial help of the banks is essential for all those companies who are in distress.

Bibliography

ALTMAN E.I., 1968. Financial Ratios, Discriminant Analysis and the prediction of Corporate Bankruptcy. *Journal of Finance*, September, pp. 589-609.

BEAVER W. H., 1966, Financial Ratios as Predictors of Failure, Empirical Research in Accounting: Selected Studies. *Journal of Accounting Research*, Supplement, pp. 71-111.

BREIMAN L., FRIEDMAN J.H., OLSHEN R.A., STONE C.J., 1984. *Classification and regression trees*, New York, Chapman and Hall.

HUARNG, K., YU, H.K., AND CHEN, C.J., 2005. The application of decision trees to forecast financial distressed companies. 2005 *International Conference on Intelligent Technologies and Applied Statistics Proceeding*, Taipei, Taiwan

OHLSON J.A. 1980. Financial ratios and the probability of bankruptcy. *Journal of Accounting Research*, Vol. 18, No. 1, pp. 109-131.

ZMIJEWSKI M.E. 1984. Methodological Issues Related to the Estimation of Financial Distress Prediction Models. *Journal of Accounting Research*, Vol. 22, pp.59-82

SUMMARY

Trigger Factors that influence Bankruptcy: a comparative and exploratory study

The phenomenon of bankruptcy has significantly influenced Italian productive environment in the last few years. The impact of the world financial crisis on the Italian economy has generated devastating effects for the wealth of the Nation, causing strong unbalances in terms of employment, productivity and investment in all sorts of industries.

The following study aims to study the financial and economic factors that have cause the failure of many firms in Italy, focusing the attention on the companies' balance sheets.

This research focuses on the situation of Lazio's companies and covers a time frame of eleven years, from 2000 to 2011 on a stratified sample of 100 firms, 50 of which are still in activity and 50 declared bankruptcy during the period 2000-2011. The attention will be focused on a cross-sectional study, considering firms at various years prior to failure.

Leonardo Di MARCO, UNINT, leonardo.dimarco88@gmail.com
Luciano NIEDDU, UNINT, l.nieddu@unint.eu

LUIGI BODIO: PROMOTER OF THE POLITICAL AND HIGH SCIENTIFIC MISSION OF STATISTICS AND PIONEER OF THE INTERNATIONAL STATISTICAL COOPERATION

Antonio Cappiello

Introduction

Luigi Bodio (October 12, 1840 Milan – November 2, 1920 Rome) was an Italian economist and statistician and one of the founders of the Italian Statistics. He was one of the 21 founding members of the International Statistical Institute (ISI) in 1885, ISI Director-General during the first 20 years (1885-1905) and ISI President for 11 years (1909-1920). A short insight into his life and his professional activity may be useful to the Italian and the International Communities interested in the history of statistics and in the origins of international statistical cooperation.

1. Early life, education and beginning of scientific activity

Luigi Bodio was born in Milan on 12 October 1840. His parents, Filippo Bodio and Giuseppina Agrati, were small business entrepreneurs. In Milan, Bodio completed his secondary education at *liceo-ginnasio Sant'Alessandro* while he pursued his university studies initially in Pavia and afterwards in Pisa where he received his Juris Doctor degree on the 7th of August 1861[1].

In 1862 he was awarded a scholarship from the Ministry of Education to attend postgraduate economics and statistics courses in Paris. In France he had important contacts with major economists and statisticians of that time, such as Louis-Adolphe Bertillon, Emile Levasseur, Antoine De Foville and Frédéric Le Play[2]. From 1864 to 1872 he taught statistics and economics in various Colleges and Professional Institutes in Livorno, Milan and Venice.

His first contributions to statistical research focused on the sources, the survey methods and objectives to be achieved in order to provide reliable scientific results to the public administration bodies and to policy makers. In particular, Bodio - as delegated expert of the Ministry of Agriculture, Industry and Commerce - produced

[1] Gilman F. H. 1910. *Luigi Bodio*. Publications of the American Statistical Association 12 (91): 283–285.

[2] Favero G. 1999. *Lo Statistico e l'industriale carteggio tra Luigi Bodio ed Alessandro Rossi (1869-1897).* Annali di Statistica serie X vol. 19 Istat. Roma; and Bonelli, F. 1969. *Luigi Bodio,* Dizionario biografico degli italiani, Istituto dell'enciclopedia Italiana, vol. 11.

an insight study (1865) on foreign trade statistics[3] and on the sources of the Italian statistical system with particular emphasis on the surveys carried out in the early years of Italian Unification[4]. Luigi Bodio's conception of statistics was grounded on rigorous scientific observations. He considered "statistics" as a main auxiliary science and as a basis for the political and economic disciplines which needed a more experimental approach. Statistics was therefore an essential instrument for measuring, not only pure economic and technical phenomena, but also social phenomena including - for instance - moral aspects of the society[5].

2. Evolution of the Italian Statistical System and Bodio's role at a national and international level

In 1872 the National "Giunta Consultiva di Statistica" (Consultancy Statistical Committee) was replaced with the "Giunta Centrale di Statistica" (Central Statistical Committee) governed by the Ministry of Agriculture, Industry and Commerce and composed of the Head of the General Statistical Directorate, the representatives of each Ministry and eight other representatives nominated in accordance with the Royal Decree of 25 February 1872 n. 708[6]. Therefore, the Central Statistical Committee became a "decisional body" as concerns statistical data to be produced by the General Statistical Directorate, and a "consultancy body" as concerns the policies of the Directorates of other Ministries. Thus, the presence of the Ministry representatives inside the Central Statistical Committee guaranteed an important connection among the activities of the various Directorates of the Ministries.

The Royal Decree 29 September 1872 n. 1048 established the separation of the Economic Directorate and the Central Statistical Directorate, the last one becoming again an autonomous division under the direct management of the Ministry. On the 30th of June 1872 Luigi Bodio was nominated Director of the "Giunta Centrale di Statistica" by Luzzatti (Ministry of Agriculture) and the 14th of November of the same year was nominated Head of the General Statistical Directorate.

After a few years, under Bodio's leadership, statistics methodology was introduced as an academic topic (Royal Decree 11 October 1875 n. 2775) and statistical courses were compulsory taught also in Law Universities. In1876 Bodio was editor, together with Cesare Correnti and Paolo Boselli, of the Archivio di

[3] Bodio L. 1865. *Saggio sul commercio esterno terrestre e marittimo del Regno d'Italia negli anni 1862 e 1863.*
[4] Bodio L. 1867. *Sui documenti statistici del Regno d'Italia*. Cenni bibliografici presentati al VI Congresso Internazionale di Statistica.
[5] Bodio L. 1869. *Della statistica nei suoi rapporti coll'economia politica e colle altre scienze affini. Prelezione al corso di statistica della Scuola superiore di commercio in Venezia.*
[6] Antonucci E. 1937. I servizi centrali, in Istat, Decennale.

Statistica" (National Statistical Archive). In the same year, Luigi Bodio conducted the first official surveys on Italian migration.

In 1885 he was founding member and Director of the International Statistical Institute (ISI) and held that post until 1905. Bodio gained a solid international reputation inside the scientific community, and his statistics methodology was considered a reference for the major academics and policy makers. Paul Leroy Beaulieu[7] claimed: "*Le statisticien qui a le plus complètement et méthodiquement réunit les documents relatifs aux mouvements de la population dans les contrées civilisées est M. Bodio le très savant chef de la statistique italienne; il les tient à jour; c' est à ses tableaux qu' il faut se reporter*".

In 1900 Luigi Bodio was elected Senator of the Italian Kingdom and from 1901 to 1904 was General Commissioner of Migration (1901-04), an inter-ministerial body created to address and protect Italian migration abroad. In 1909 he was elected President of the International Statistical Institute and held that post until his death in 1920.

Certainly Bodio gave an enormous prestige to Italian statistics,[8] but his main contribution was especially to "statistics beyond national and local borders". This was well recognized, also after Bodio's death, by the international statistical community. Friedrich Zahn, President of the International Statistical Institute from 1931 to 1936, said that Bodio "*appartient avec Quetelet, Engel, Lavasseur, Bertillon, von Ottingen, von Mayr, Wappäus, von Neumann-Spallart et autres à ce noyau d'hommes de science et d'administration éminents, qui firent parvenir la statistique à une âge d'or dans la seconde moitié du dix-neuvième siècle*[9]".

Bodio was convinced of the high political and scientific role of statistics. He enriched Italian official statistics with new methodological concepts and improved terminology, and obtained an important international reputation coordinating the activities of the International Statistical Institute. Consequently, "statistics" became a reputable "institution"[10] in the Italian "state apparatus", a discipline - as previously said - taught in most universities and high schools[11].

[7] Revue des deux mondes, tome 143, 15 Octobre 1897, p.864.

[8] Leti G. 2004. *The International Activities of Italian Statisticians Prior to the Second World War*. Statistica, anno LXIV, n. 2.

[9] Trad. : "Bodio belongs, together with Quetelet, Engel, Lavasseur, Bertillon, von Ottingen, von Mayr, Wappäus, von Neumann-Spallart and others, to the new eminent experts of Science and Administration who conducted statistics to a golden age in the second half of the XIX century".

[10] According to Bodio "statistics and economics should necessarily prepare and follow the legislation work".

[11] Patriarca S. 2003. *Numbers and Nationhood*. Cambridge University Press.

3. Reorganization of official statistics in Italy and knowledge sharing with leading economists and statisticians

The activities of the General Directorate of Statistics[12] under the guidance of Luigi Bodio, as before mentioned, represented a prosperous period for Italian statistics that for many years was among the most advanced in Europe[13].

Bodio gave to the Central and local Statistical Offices a very efficient organization. He strengthened the power of the Central Office allowing it to make use of all the State Administration Offices for collecting statistical information and formulating appropriate guidelines on programs and methods of the State Administration statistical procedures.

He also centralized the analysis of all national statistics at the General Statistical Directorate (Central Statistical Office). Therefore the responsibilities of the General Statistical Directorate were expanded and covered almost all the economic, demographic, social and political fields.

Bodio's efforts in building official statistics were directed toward the recognition of the State Administration's own role, independent from political influences as veritable bureaucratic power. For this reason, Bodio promoted an "interpretative prudence" which emphasised the technical role of statistics[14]. Bodio's personal international prestige gave continuity to the application of the scientific methodology to Italian statistics.

From a practical point of view, in order to speed up the data analysis, Bodio proposed to automate data processing by adopting classification machines based on perforated files invented by Herman Hollerith and used in United States for the census in 1890. In 1901 Bodio's initiative allowed, for the first time in Italy, a partial automation of data processing.

Luigi Bodio, as previously stated, made a great contribution to methodological and applied statistics at the international level. He analysed issues concerning the pure statistical methodology such as sample representativeness proposed by Kiær[15], as well as applied statistics. As concerns applicative aspects of statistical data, he dedicated particular attention to crime statistics, health statistics, educational statistics and some specific aspects of migration, the natural movement of

[12] National (Central) Statistical Office.

[13] Evidence of Bodio's scientific activity are also represented by the numerous scientific reports attached to the official statistics of the Italian Kingdom (1872-1900), the proceedings of the Supreme Council of Statistics (published in the *Annali di Statistica*), of the State Council, of the Commissioner for the emigration, and from the private and official correspondence with politicians, statisticians and economists. We do not have exhaustive references of his scientific work since most part of Bodio's archive in the General Directorate of Statistics was lost. Bodio's private library was donated to the Bocconi University of Milan where is currently stored.

[14] Favero G. 2006. I servizi statistici ufficiali in Italia dall'Unità alla Repubblica: strategie di organizzazione interna e pertinenza dell'informazione prodotta. Ca' Foscari di Venezia No. 02/NL/2006.

[15] Anders Nicolai Kiær (1838–1919) was a Norwegian statistician who first proposed the method of the representative sample in order to obtain information on a population.

population and the economic indexes. Luigi Bodio precisely defined some sub-categories of the migration phenomena and focused on aspects such as periodic, seasonal and temporary migration, and permanent migration (classifying the data on the basis of the information concerning the migrants' residence abroad[16]).

As a means of knowledge sharing about methodological and applied statistics, Luigi Bodio often used handwritten letters, the most employed instrument of communication of that period. Bodio's letters to economists, statisticians, intellectuals and politicians constitute a veritable treasure of wisdom composed of scientific and political reflections and advice. For instance, the letters of his correspondence with the economists Vilfredo Pareto and Edwin Robert Anderson Seligman contain very valuable sources of knowledge of the economic and political thought of that time.

In 1920 Luigi Bodio chaired the works of the International Statistical Commission appointed by the Council of the League of Nations[17]. He died in Rome on the 2nd of November 1920 at the age of 80 years.

In 1996, a Statistical Cooperation Association (named *ICstat, International Cooperation Center for Statistics "Luigi Bodio"*) was dedicated to the memory of Bodio in recognition of his dedication and promotion of the statistical cooperation. The Association, based in Rome, promotes international cooperation in the field of statistics, economics and law. ICstat coordinates technical assistance projects financed by international institutions and is particularly involved in several transition and developing countries. The association supports democratic governance, crisis prevention and recovery, human rights application and monitoring systems, post-conflict political elections and referendum. Moreover, ICstat produces studies for scientific and policy purposes.

Acknowledgments

I would like to acknowledge the contribution to international statistical cooperation given - following Bodio's spirit - by Dr. Salvatore Favazza, Senior Statistician, Officer at United Nations and former Director of *ICstat, International Cooperation Center for Statistics "Luigi Bodio"*.

[16] Ratti A. M. 1929. *Italian Migration Movements, 1876 to 1926.* International Migrations, Vol. II, edited by Walter F. Willcox, 440- 470, New York.

[17] Coats R. H. 1921. *Report of the International Statistical Commission Appointed by the Council of the League of Nations.* Quarterly Publications of the American Statistical Association, Vol. 17, No. 133 (Mar.), pp. 635-638.

References and further readings

ASCHIERI A. 1921. *Luigi Bodio*. *Metron*, I (21), n. 4, pp. 180-195.

BENINI R. 1925. *Atti della Giunta Centrale di Statistica* 11/1920. *Annali di Statistica* Volume XI, n.5.

BONELLI F. 1969. *Luigi Bodio.* Dizionario biografico degli italiani, Istituto dell'Enciclopedia Italiana, volume XI.

COATS R. H. 1921. *Report of the International Statistical Commission appointed by the Council of the League of Nations*. Quarterly Publications ASA, Vol. XVII, n. 133.

FAVERO G. 1999. *Lo Statistico e l'industriale carteggio tra Luigi Bodio e Alessandro Rossi (1869-1897).* Annali di Statistica serie X vol. XIX, Istat. Roma.

FAVERO G. 2006. *I servizi statistici ufficiali in Italia dall'Unità alla Repubblica: strategie di organizzazione interna e pertinenza dell'informazione prodotta*. Nota di Lavoro del Dipartimento di Scienze Economiche Università Ca' Foscari di Venezia No. 02/NL/2006.

GILMAN F. H. 1910. *Luigi Bodio*. American Statistical Association, Volume XII, n. 91.

LETI G. 2004. *The International Activities of Italian Statisticians Prior to the Second World War*. Statistica, anno LXIV, n. 2.

MAYOR DES PLANCHES E. 1921. *Bollettino dell'Emigrazione*. *Nuova Antologia, volume* XX, *n. 16*

NIXON S. W. 1960. *A history of ISI, 1885-1960.* The Hague, pp. 15-31.

PATRIARCA S. 2003. *Numbers and Nationhood*. Cambridge University Press.

RATTI A. M. 1929. *Italian Migration Movements, 1876 to 1926.* International Migrations, Vol. II, edited by Walter F. Willcox, pp. 440- 470, New York.

REVUE DES DEUX MONDES, Tome 143, LXVIIè année, 4è période, 15 Octobre 1897.

STRINGHER B. 1920. *Rendiconti dell'Accademia Nazionale dei Lincei.* Sess. 5, XXIX.

SUMMARY

Luigi Bodio: promoter of the political and high scientific mission of statistics and pioneer of the international statistical cooperation

Luigi Bodio (1840-1920) was an Italian economist and statistician, considered one of the founders of the Italian Statistics. He was one of the 21 founding members of the International Statistical Institute (ISI) in 1885, ISI Director-General during the first 20 years (1885-1905) and ISI President for 11 years (1909-1920). This article gives a short insight into Luigi Bodio's life and may be useful to the Italian and the International Communities interested in the history of statistics and in the origins of international statistical cooperation.

Antonio CAPPIELLO, Senior Economist, economics@live.co.uk

SOCIETÀ E RIVISTA ADERENTI AL SISTEMA ISDS
ISSN ASSEGNATO: 0035-6832

Direttore Responsabile: Dott. CLAUDIO CECCARELLI

Iscrizione della Rivista al Tribunale di Roma del 5 dicembre 1950 N. 1864

Associazione all'Unione Stampa Periodica Italiana

TRIMESTRALE

La copertina è stata ideata e realizzata da Pardini, Apostoli, Maggi p.a.m.@tin.it – *Roma*

Stampato da CLEUP sc
"Coop. Libraria Editrice Università di Padova"
Via G. Belzoni, 118/3 – Padova (Tel. 049/650261)
www.cleup.it