rivista di statistica ufficiale

REVIEW OF OFFICIAL STATISTICS

n. 1-2-3
2023

In this issue:

Istat
Istituto Nazionale di Statistica

rivista di statistica ufficiale

REVIEW OF OFFICIAL STATISTICS

n. 1-2-3
2023

In this issue:

Istat

Editor:
Patrizia Cacioli

Scientific committee

President:
Gian Carlo Blangiardo

Members:
Corrado Bonifazi
Vittoria Buratta
Ray Chambers
Francesco Maria Chelli
Daniela Cocchi
Giovanni Corrao
Sandro Cruciani
Luca De Benedictis
Gustavo De Santis
Luigi Fabbris
Piero Demetrio Falorsi
Patrizia Farina
Jean-Paul Fitoussi
Maurizio Franzini
Saverio Gazzelloni
Giorgia Giovannetti
Maurizio Lenzerini
Vincenzo Lo Moro
Stefano Menghinello
Roberto Monducci
Gian Paolo Oneto
Roberta Pace
Alessandra Petrucci
Monica Pratesi
Michele Raitano
Maria Giovanna Ranalli
Aldo Rosano
Laura Terzera
Li-Chun Zhang

Editorial board

Coordinator:
Nadia Mignolli

Members:
Ciro Baldi
Patrizia Balzano
Federico Benassi
Giancarlo Bruno
Tania Cappadozzi
Anna Maria Cecchini
Annalisa Cicerchia
Patrizia Collesi
Roberto Colotti
Stefano Costa
Valeria De Martino
Roberta De Santis
Alessandro Faramondi
Francesca Ferrante
Maria Teresa Fiocca
Romina Fraboni
Luisa Franconi
Antonella Guarneri
Anita Guelfi
Fabio Lipizzi
Filippo Moauro
Filippo Oropallo
Alessandro Pallara
Laura Peci
Federica Pintaldi
Maria Rosaria Prisco
Francesca Scambia
Mauro Scanu
Isabella Siciliani
Marina Signore
Francesca Tiero
Angelica Tudini
Francesca Vannucchi
Claudio Vicarelli
Anna Villa

Editorial support: Alfredina Della Branca

rivista di statistica ufficiale
n. 1-2-3/2023

Four-monthly Journal: registered at the Court of Rome, Italy (N. 339/2007 of 19th July 2007).

e-ISSN 1972-4829
p-ISSN 1828-1982

© 2023
Istituto nazionale di statistica
Via Cesare Balbo, 16 – Roma

Unless otherwise stated, content on this website
is licensed under a Creative Commons License - Attribution - 4.0.
https://creativecommons.org/licenses/by/4.0/deed.it
Data and analysis from the Italian National Institute of Statistics can be copied,
distributed, transmitted and freely adapted, even for commercial purposes,
provided that the source is acknowledged.
No permission is necessary to hyperlink to pages on this website. Images, logos
(including Istat logo), trademarks and other content owned by third parties belong
to their respective owners and cannot be reproduced without their consent.

The views and opinions expressed are those of the authors and do not necessarily reflect the official policy or position of the Italian National Institute of Statistics - Istat.

The Scientific Committee, the Editorial Board and the authors would like to thank the anonymous reviewers (at least two for each article, on a voluntary basis and free of charge, with a double-anonymised approach) for their comments and suggestions, which enhanced the quality of this issue of the Rivista di statistica ufficiale.

Expected developments in the number and type of Italian households until 2040

Martina Lo Conte[1], Marco Marsili[1], Gianni Corsetti[1], Eleonora Meli[1]

Abstract

Household projections provide crucial information for policy planning. The purpose of this work is to estimate household projections for Italy, consistently with the official population projections, annually updated by the Italian National Institute of Statistics (Istat). An approach based on the Propensity rate model met the requirements: parsimony, simplicity, replicability and quality. The method, in addition to producing the number of households by type, provides with future time series of the population by household position (child, living with a partner with or without children, lone parent, living alone, other position), age and sex. Household projections from 2020 to 2040 were released in 2021 for the first time together with population projections. The results show an increase in the number of households by one million and a decrease in their average size, which would drop from 2.3 members to 2.1 Furthermore, the results document a decrease in couples with children, an increase in those without children and of people living alone.

Keywords: Family, household, Italy, official household projections, propensity rate.

DOI: https://doi.org/10.1481/ISTATRIVISTASTATISTICAUFFICIALE_1-2-3.2023.01.

1 Martina Lo Conte (loconte@istat.it); Marco Marsili (marsili@istat.it); Gianni Corsetti (giacorsetti@istat.it); Eleonora Meli (meli@istat.it), Italian National Institute of Statistics - Istat.

The views and opinions expressed are those of the authors and do not necessarily reflect the official policy or position of the Italian National Institute of Statistics - Istat.

The authors would like to thank the anonymous reviewers for their comments and suggestions, which enhanced the quality of this article.

1. Introduction

Literature has shown the importance of the family dimension in the study of many social and economic phenomena, such as housing, health, work, welfare, migration, poverty and social exclusion (Bell *et al.* 1995; Hantrais *et al*. 2006; Hays 2002; Paciorek 2013). Family formation, in fact, directly or indirectly influences most social and economic outcomes.

Since the 1960s, families in many Western countries have undergone significant changes. The extended family has almost disappeared and the traditional two-parent family has become less common, as divorce rates, remarriage and single-parenthood have increased. Families have seen more women enter the labour market, teenagers spend more time in education and older family members live longer and, increasingly, alone (Stevens et al. 2012).

Data on the number of households, their typology and the distribution on the territory, today and in the years to come, represents a central support for planning public intervention actions. Policies supporting fragile families, particularly older adults living alone, single-parent families, or large families can take advantage of this kind of information.

Household projections are also useful for urban and housing policies, as well as for estimating energy consumption. Notably, a growing number of households, linked to a reduction in their size, may have significant effects on energy use (O'Neil *et al.* 2002).

In addition, the development of predictive scenarios for population, households, and families meets a primary knowledge need for countries that, like Italy, have undergone profound demographic and social transformations over the years.

In 2020, the Italian National Institute of Statistics (Istat) started a project on household projections integrated with the established set of official population projections. The goal of the project was to find a model to estimate the number and type of households in Italy over the next 20 years, which met the following requirements: consistency with the demographic projections, parsimony (detailed and quality information versus cost), need for non-burdensome input data, timeliness, and annual reproducibility.

This paper presents the model used to project households in Italy from 2020 to 2040 and reports the main results. It is structured as follows. Section 2 gives an illustration of the most relevant household projection models available in the literature. The chosen model and strategy are presented in Section 3. The data and definitions used in this study are discussed in Section 4, followed by the step-by-step application in Section 5. Section 6 shows the results obtained. Finally, some conclusions and possible next steps are drawn up in Section 7.

2. Models in literature

There is a considerable availability of models for household projection purposes. Following the classification proposed by Gill and Keilman (1990), household forecasting models can be distinguished according to the approach followed (static or dynamic) and/or to the basic unit of analysis (micro or macro)[2].

In a static approach, projections are carried out applying proportions or rates that allow shifting from a projected population to the corresponding families and households. The best known is the “headship rate method” (United Nations 1973; Kono 1987; Linke 1988), widely applied for its reproducibility and the simplicity of the procedure. This model focusses on the characteristics of the household head, producing outcomes about the total number of households and their mean size.

These limitations (see next paragraph) are partly overcome by another static model, the propensity rate model, which offers a wider range of results, such as the distribution of households by type and population by family position (de Beer *et al*. 1999; Wilson 2013; ABS 2019).

Dynamic approaches, unlike static ones, explicitly model family events. In the dynamic method, the distribution of the population broken down by state is the outcome of transitions people make in their lives. Individuals move between states and consequently, the structure of the households change (multistate method). Therefore, dynamic models give a more accurate picture of the mechanism of social and demographic changes occurring in the real world. As a result, they are in theory better suited for integrating population projections with household projections. In addition, the study of transitions allows assessing the impact on population dynamics caused by policies or by socio-economic changes that may emerge in the society.

Depending on the basic unit used, dynamic models can be defined as macro or micro (Keilman *et al.* 1988). Macro-dynamic models proceed by groups based on different characteristics, which vary according to the application (*e.g.* age, sex, marital status, family position, etc.). Rates for demographic events and household formation and dissolution are then applied to the multi-

2 For an in-depth review on household forecasting methods, see Keilman (2019).

group population. This allows modelling transitions between states for each calendar year along the time horizon. The Lipro macro-dynamic model, for example, implements a multi-state demographic model that focusses on flows between states (*e.g.* from *living alone to living in a couple* or from being a *partner with children* to being a *single parent*) (van Imhoff *et al*. 1991). Similarly, *ProFamy* provides detailed family compositions and population group characteristics such as marital/union status. It uses socio-demographic rates as input, projecting family states of individuals grouped by age-cohort and other attributes (Zeng *et al*. 2013).

Starting from the individual's risk of experiencing a specific event, dynamic micro-based models consider the individual as the basic unit of prediction, with the aim of simulating its entire life cycle (Cannari *et al.* 1998). The dynamics of the population are derived by aggregating the forecast results according to the chosen classificatory criteria. Microsimulation modelling is a very useful tool for measuring the impact of certain changes, such as the effect of ageing on the pension system, or of policies, such as the impact of child support on female labour force participation and household income (Décarie 2012).

To overcome some limitations that emerge in the two different approaches, there have been some attempts to use the static and dynamic models in a combined way. For example, de Beer and Alders (1999) developed a probabilistic forecasting model, which first projects the population by marital status based on a multistate model and then applies propensity rates to obtain households. Similarly, Mic/Mac represents an attempt to bridge the gap between micro and macro models (Willekens *et al.* 2007), a methodology that complements conventional projections by age and sex (aggregate projections of cohorts, Mac) with projections of the way people live their lives (projections of individual cohort members, Mic). In Mic/Mac the life course is viewed as a sequence of states and events that result in transitions from one state to another. An advantage of such an approach is a better control for population heterogeneity, while traditional projections assume that members of a cohort are identical with respect to socio-demographic behaviour. Another benefit offered by Mic/Mac is the information on duration of stay, generally not available in traditional models without a life course perspective. Lastly, the projection of people's lives is taken under control as Mac can be used as a starting point for Mic (validation/calibration of Mic by using Mac).

Forecasting always involves dealing with uncertainty, since the future is inherently uncertain. Therefore, for each of the above approaches, a further distinction between probabilistic and deterministic models can be made. The former model tells us how likely it is that the number of households for a given future year will be within a certain range; the latter one predicts one result, the most likely trajectory, sometimes formulating alternative (usually low and high) scenarios (Keilman 2019).

3. The choice of the method

The main objective of the project was to find a projection model that would make it possible to jointly release official population and household projections on an annual basis. Therefore, we required a model that was consistent with demographic projections, parsimonious (with detailed, high-quality results respect to costs), and replicable every year. In this regard, an important aspect to analyse concerned the availability, but above all, the timeliness, of the data sources to be used as input for our model. The continuous working cycle, due to the current production of the official basic data (census, population dynamics and social surveys), preparatory to the construction of a forecasting model on households, requires a step-by-step process that in Istat necessarily has to be completed by the end of the following year to which the information refers. Therefore, although data availability and timeliness should not drive the choice of an optimal model, for a national statistical institute, such as Istat, this factor cannot be ignored.

The previous considerations about time constraints, along with the desire to refresh the assumptions underlying the projections on a yearly basis, limit our potential range of action on the methodological side, particularly with regard to the dynamic models and, above all, the micro-dynamic ones. Dynamic models certainly allow a more realistic representation of the population development due to demographic and social processes (birth, death, marriage, divorce, migration events, etc.). However, such an approach also presents some drawbacks from the perspective of a national statistics institute. Among them, it has to be stressed that micro-dynamic models require a huge amount of data processing, to the point that generally only a small representative sample of the population can be processed in the microsimulation procedure. Secondly, the final results are based on a collective of individual trajectories not necessarily bound to any main result at the macro level, so such an approach may require the adoption of considerable calibration measures in order to obtain consistent predictions.

As mentioned in the previous paragraph, Mic/Mac represents a valid solution for overcoming the problem of potential inconsistencies in the results at the micro level. Nevertheless, Mic/Mac has been designed to be used mainly for single countries separately. The model is substantially a uni-regional type, therefore not particularly suitable for a country like Italy, where the multiregional dimension is essential in explaining social and demographic behaviours.

Because of the low detailed outcome, another approach not able to best represent the complexity of the Italian reality is the headship rate model. This method includes only the characteristics of the household head, which is often defined vaguely and differently across countries (Murphy 1991).

In conclusion, for the above-mentioned reasons and because of its adaptability to the Italian situation, the method that best met our requirements is the static approach based on the “Propensity rate model”. This method has been used in recent years by the Australian Bureau of Statistics to project households in Australia and New Zealand (ABS 2019). It goes beyond the classical headship rate model, overcoming the concept of ‘headship’ and providing a more detailed set of information (McDonald *et al.* 2006; Bell *et al.* 1995; Wilson 2013; Blangiardo *et al*. 2012).

The model relies on propensity rates, defined as the proportion of people of a certain age and a specific household position in a given year. As an example, the propensity of a 30-year-old person to live in a couple with a partner is the ratio of the number of 30-year-olds living in a couple to the total 30-year-old population.

Starting from a predicted population by age, sex and region, the method develops in five steps to predict the population by household position; this allows the number and type of households to be determined. The steps, described in detail in Section 5, are the following:

Step 1. Estimating the projected population living in households.

Step 2. Calculating household propensity rates.

Step 3. Modelling the future trends of household propensity rates.

Step 4. Obtaining the projected population by household position.

Step 5. Calculating the number, type, and size of the projected households.

Generally, the propensities are extrapolated from the analysis of past trends. Different methods can be applied, such as linear or non-linear models, or even following experts’ judgments. In this application, we introduced a new indicator that measures life expectancy in a specific household position, to be used for predicting future household behaviour (see Section 5).

There are several advantages of the method: it easily links to existing population projections; data inputs are not as onerous as for dynamic models; it provides detailed results. In light of this, projections can be easily updated when new input data become available.

However, there are also some drawbacks, which mainly derive from the static nature of the method. In fact, the application of propensity rates to the population may lead to inconsistencies in terms of overall results. For example, the projected number of male and female partners may differ (the two-sex problem, Keilman 1985; Schoen 1988) or the initially projected propensities will not sum to unity across living arrangements, and therefore, require ex-post adjustments (Wilson 2013).

4. Data and definitions

We integrated several data sources to carry out our analysis. Firstly, the Permanent Population and Housing Census supplied the base-population on January 1st 2020 by age, sex, region of residence, and type of residence (private or institutional household[3]). Since 2018, the new population and housing census, which replaces the previous one traditionally based on a decennial basis, releases data annually through the integration of the information available from administrative sources and that were acquired with sample surveys. The combined use of data from registers and surveys guarantees the information on the main demographic and socio-economic characteristics of the resident population in Italy. The specific objective of the permanent census, in particular, is the production of data relating to the counting of the resident population at the municipal level and its distribution by sex, age, citizenship, level of education, and professional status (Istat 2021*c*).

Secondly, information on household and family structures was achieved from the Italian multipurpose survey *Aspetti della vita quotidiana*, which provided a long time series (from 2002 to 2019). This is an annual cross-section sample survey carried out by interviewing a sample of 20,000 households (for a total of about 50,000 people) and guarantees consistent and accurate estimates at the regional level. The questionnaire includes a detailed set of questions regarding the family context of the individuals, so that the information on households is the strength of the survey, and represents the benchmark for the social statistics produced by Istat (Bagatta *et al.* 2006).

Both the Permanent Population Census and the Italian multipurpose survey are based on the population usually resident[4] in Italy. Regarding the population living in households, both sources rely on the definition of the *de facto* situation.

3 An institutional household is a group of people cohabiting for caring, military, punitive, religious or other similar reasons, and thus residing in institutions such as hospitals, barracks, prisons, nursing homes or religious buildings.

4 According to the "Regulation (EU) N. 1260/2013 of the European Parliament and of the Council on European demographic statistics", the definition 'usual residence' means the place where a person normally spends the daily period of rest, regardless of temporary absences for purposes of recreation, holidays, visits to friends and relatives, business, medical treatment or religious pilgrimage.

Regarding the institutional households, for the Multipurpose survey, they are not part of the reference population. For the Census, the population resident in institutions refers to the administrative information (Population Register).

In the model, we use the following definitions of family and household. People in a couple or a parent-child relationship usually living together (see footnote 4) form a family (or nucleus). This includes a married, civilly united (same sex) or cohabiting (both same and opposite sex) couple, with or without children, or a single parent with one or more children. Children refer to a never-married biological, step/adopted son or daughter (regardless of age), who live with at least one of the parents, and who have no partner or own children in the same house. In case of separated parents, minor children are considered part of the parent's family to whom they have been assigned by the court; if children are adults, they can decide which parent to join.

The household may consist of one person (one-person household) or a group of people. In the latter case, two conditions are required:

1. co-residence;
2. the presence of a relationship such as marriage, civil union, kinship, consensual union, friendship, adoption, guardianship, or emotional ties. Conversely, guests, servants or persons who share the dwelling for economic reasons (tenants, boarders, etc.) are not considered as members of the household.

A household may contain several families (family-households). In this case, for instance, a child who marries, if he/she continues to live with his/her parents constitutes a new family within the parents' household. Also, a daughter-in-law who lives with her in-laws forms a household, even in the absence of her husband (being linked to them by an emotional tie).

In addition, a household may consist of people who live together without forming any family, but may be related (such as two siblings or cousins, friends or parents with a divorced/widowed child): these are the so-called multi-person households.

In this research, households containing two or more families will be considered together, as they represent a small share of the total number of households (about 1.5 per cent in the last two decades). Table 4.1 shows the eight household positions and six household types considered.

Table 4.1 - Household type and household position classifications

Household type	Household position
1. One-person household	1. Lone person
2. Couple without children (a)	2. Partner in couple with children
3. Couple with children (a)	3. Partner in couple without children
4. One parent with children	4. Single parent with children
5. Multi-person household	5. Child (b)
6. Two or more family household	6. Other person in one-family households (c)
	7. Person living with others not forming a family
	8. Person in a household with 2 or more families

Source: Authors' processing
(a) Both married and unmarried couples.
(b) Children are never married biological, step/adopted son or daughter (regardless of age), living with at least one parent in a couple or with a single parent, and who have no partner or own children in the same household.
(c) Other person is a person living in a one-family household, not having a couple relationship or parent-child relationship, such as, for example, a cousin or a friend.

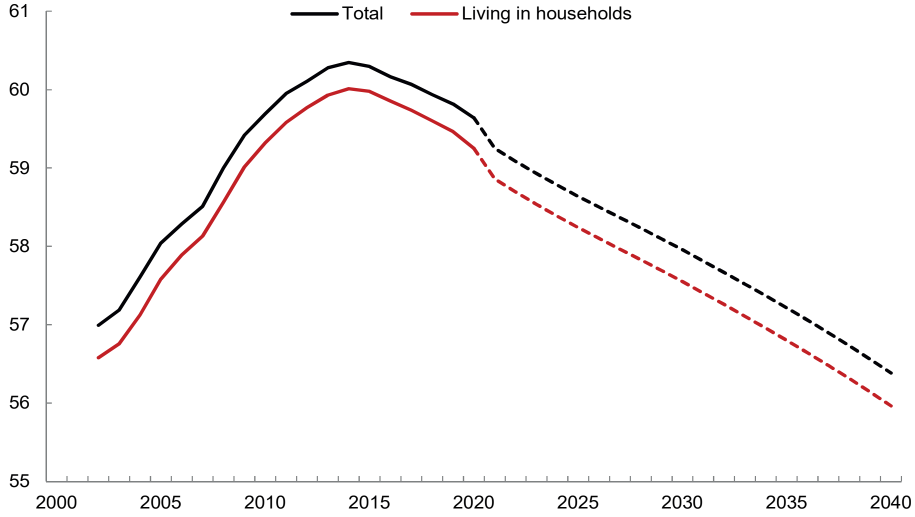
5. Application of the method

In the first step, the aim is to estimate the projected population living in private households. In fact, since the base population, as well as the projected one, represents the entire population, it is necessary to separate individuals living in private households from those living in institutions such as hospitals, barracks, prisons, nursing homes, religious buildings, etc. (see footnote 4).

The proportions of people living in institutions by gender, five-year age group and region were calculated using data from the Permanent Census (mean value observed in 2019). In general, such proportions present a substantial stability along time, despite the ongoing and strong ageing process of the country. In the period 2002-2019 the total amount of people living in institutional households ranges from about 310 thousand to about 480 thousand, thus representing from 0.5% to 0.8% of the total resident population. Given their substantial stability, the proportions of people living in institutional households by sex, age group and region have been kept constant along the time horizon. Among the motivations for assuming a constant trend for the future, we find particularly relevant the fact that older people (the prevailing component in this particular sub-population) are expected to achieve a longer healthy life in the years to come (Quattrociocchi *et al*. 2021). Moreover, family support is still very strong in Italy, where older people are often cared for by family members or, at their own home, by caregivers. We then assumed that both the improved quality of life in old age and the family support could compensate for the large increase of the older population in the future. Therefore, we chose a conservative scenario, following the current trends with a constant proportion of people living in institutional households by age, gender, and region.

Once institutional households are left out of the analysis, we obtain the population living in private households from 2020 to 2040 (Figure 5.1). The graph also shows the growth of the Italian population at the beginning of the millennium, which reached around 60 million between 2012 and 2017, and then began to decline. In fact, a sharp decline in births and an increase in deaths started in the early 1990s, leading to a negative natural balance. For example, in 2020 there were about 400,000 births and over 700,000 deaths. Since 2015, the migration balance has no longer compensated for the natural balance, resulting in a negative total balance that is leading to a continuous population decline.

Figure 5.1 - Total population and population living in households. Italy, years 2002-2040 (observed and median scenario, in millions)



Source: Authors' processing on Population Census and Population projections

The second step of the model consists in calculating the propensities to live in the household positions of interest. For each time t, a propensity rate is defined as the proportion of persons of age x and sex s with a household position i:

$$\textbf{Propensity Rate}_{x,i,s,t} = \frac{P_{x,i,s,t}}{P_{x,s,t}} = \textbf{LAP}_{x,i,s,t}$$

where x= five-year age group 0-4, 5-9, ..., 80-84, 85+,
s=sex,
i=household position,
t=time.

Hereinafter, these rates are also referred to as *Living Arrangement Propensities* (LAP). Survey data from *Aspetti della vita quotidiana* (AVQ) allows computing LAPs over the 20-year period from 2002 to 2019. These rates are used to break down the projected population by living arrangement at the regional detail. However, due to the sparseness of data in small regions, we decided to aggregate the LAPs, working first at the 'macro-region' level.

In order to identify homogeneous groups of regions, characterised by similar family structures and evolution over time, we ran a dynamic principal component analysis using the STATIS methodology (Lavit *et al*. 1994). The analysis examined the main socio-demographic variables at the regional level over the period 2002-2019, including fertility rates, mean age at childbearing, mean household size, separation and divorce rates, female employment rates, internal and international migration rates, and proportions of certain family types (singles, couples with and without children, lone parents, etc.). The procedure was optimised by eliminating co-variates with low latent variability explained by the axis. Finally, we obtained the following five clusters:

- Group 1 - North-west (Piemonte, Valle d'Aosta/V*allée d'Aoste*, Lombardia, Liguria);
- Group 2 - Eastern Adriatic (Veneto, Emilia-Romagna, Trentino-Alto Adige/*Südtirol*, Friuli-Venezia Giulia, Marche);
- Group 3 - Tyrrhenian (Toscana, Lazio);
- Group 4 - South (Campania, Puglia, Calabria, Sicilia);
- Group 5 - Central (Umbria, Sardegna, Abruzzo, Molise, Basilicata).

Such a geographical breakdown reflects both the proximity of regions and their similarity in a socio-demographic perspective. The quality of the results allowed clustering the smallest regions in an effective way.

In the third step, the aim is to model future household propensity rates. In this regard, we introduce a synthetic indicator calculated as the sum of the propensity rates by age. In particular, we use the "number of years lived" function from the life table (L_x) to weight the propensity rates, as proposed in the Sullivan's methodology (Sullivan 1971). We denote this indicator as the *Total Household Position Rate* (TPR):

$$\mathbf{TPR_{i,s,t}} = \sum_{x=0-4}^{85+} \mathbf{LAP_{x,i,s,t}} * \mathbf{L_{x,s,t}} = \sum_{x=0-4}^{85+} \frac{\mathbf{P_{x,i,s,t}}}{\mathbf{P_{x,s,t}}} * \mathbf{100} * \mathbf{L_{x,s,t}}$$

where i=household position, s=sex, x=five-year age class, t=time from 2002 to 2019. $L_{x,s,t}$ represents the number of years lived in age group x during year t by individuals of sex s; the indicator is collected from the official Life tables in all the years from 2002 to 2019[5].

5 Source: Istat, *Life tables of the resident population*, https://demo.istat.it/app/?i=TVM&a=1974&l=en.

Under the hypothesis of independence between mortality and household position, the TPR for a given household position would represent approximately how many years on average a cohort of individuals will expect to live in that position. Such hypothesis implies that along the life course the family behaviours and the mortality conditions are experienced as observed in a given calendar year. Furthermore, because of population heterogeneity, people living in different household positions (for example living alone vs. living in couple) also present different mortality risks. However, we assume that the error made by attributing the same mortality to different family types has a limited impact. In fact, in old age, where mortality is higher, the prevailing family condition is living alone. On the contrary, in adulthood, when there is greater heterogeneity between family positions, the risk of death is low and therefore the impact of the error is rather negligible.

In conclusion, the TPR indicator, despite the limitations described above, allows the construction of forecast assumptions that make logical sense and that can be kept under control.

As shown in Table 5.1, the mean time spent as a single person has increased substantially: while in 2002, a man counted on living an average of 5.8 years as a single person (out of a total life expectancy of 77.2), in 2019 the estimated time in this state rises to 9.4 years (out of 81). In contrast, due to declining birth rates, in 2002, women expected to live 22.7 years in a couple with children (out of a total of 83), but in 2019 this expected time drops to 19.6 years (out of a total life expectancy that in the meantime has risen to 85.3 years). In addition, the time in 'child' status has increased from 30.4 to 31 years for males and from 27.7 to 28.6 for females. This is due to the typical Italian behaviour of young people prolonging their stay in the family of origin[6] (Castagnaro *et al*. 2022).

In order to model future trends of propensities, we decided for a top-down approach where we first project the TPR per household position and macro-region, then we estimate the age pattern for each projected year ($LAP_{x,i,s,t}$). In fact, one limitation of the static approach based on propensity rates is the distinct prediction of the rates by single age group, as it is difficult to control their consistency, particularly in the patterns of small regions, with the risk of obtaining unreliable results. Predicting the TPR first made it easier to translate assumptions about family behaviour and to hold together future trends in the various household positions.

6 Children are considered as such if they are never married, regardless of age.

Table 5.1 - Total household position rates by family position and sex. Italy, years 2002-2019

Household position	Men					Women				
	2002	2005	2010	2015	2019	2002	2005	2010	2015	2019
Lone person	5.8	6.1	7.5	8.5	9.4	10.7	11.1	12.0	12.5	12.9
Person in multi-person household	0.9	1.1	1.1	1.4	1.6	1.7	1.7	1.6	1.7	1.7
Partner without children	13.2	13.6	14.7	14.1	13.7	12.2	12.6	13.7	13.2	12.7
Partner with children	23.2	22.6	21.6	20.4	19.7	22.7	22.0	21.1	20.2	19.6
Lone parent	0.8	1.1	1.0	1.2	1.4	4.2	4.3	4.6	4.8	5.4
Child	30.4	30.7	30.2	30.9	31.0	27.7	28.1	27.8	28.3	28.6
Other position	0.8	0.6	0.8	0.9	0.9	1.5	1.4	1.1	0.9	1.0
Person in household with 2+ families	2.1	2.3	2.4	2.7	3.3	2.3	2.3	2.4	3.0	3.4
Total	**77.2**	**78.1**	**79.3**	**80.1**	**81.0**	**83.0**	**83.5**	**84.3**	**84.6**	**85.3**

Source: Authors' processing on "Aspetti della vita quotidiana" survey data

The total time spent in each household position ($TPR_{i,s,t}$) by macro-region is predicted using time-series analysis models based on trend extrapolation from the 2002-2019 period (Box *et al*. 2015). For each household position and sex, the predictions have been carried out applying a best ARIMA optimisation procedure as shown in Table 5.2. These models proved to be effective for all five macro-regions.

Table 5.2 - Predictive models of Total household position rates by position and sex (a)

Household position	Men	Women
Lone person	RWD || ARIMA (1,0,0)	RWD
Person in multi-person household	RWD	RWD
Partner without children	ARIMA (2,0,0)	RWD
Partner with children	RWD || ARIMA (2,1,0)	RWD || ARIMA (2,1,0)
Lone parent	RWD	RWD || ARIMA (2,0,0)
Child	RWD	RWD
Other position	RWD	ARIMA (1,0,0)
Person in household with 2+ families	ARIMA (1,1,0)	ARIMA (1,1,0)

Source: Authors' processing
(a) RWD=Random Walk with Drift model; ARIMA=Auto Regressive Integrated Moving Average model.

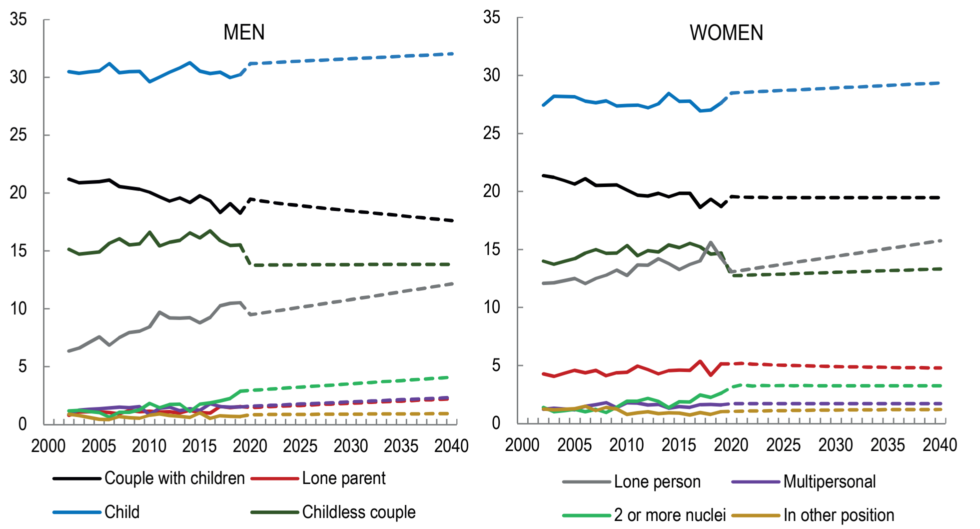
The expected changes in the time spent in different household positions over the 20-year projection period reflect the assumptions underlying our projections. In summary, they show:

- an increase in "lone persons";
- a fall in "partners with children";
- a slight growth in "partners without children";

- an increment in the “child” position;
- a small rise in “lone parents”, especially fathers;
- a substantial stability of “other people” living with a nucleus;
- a slight increase in “persons in households with 2 or more families”.

These trends can be observed in Figure 5.2, which shows, as an example, the results for the North-west macro-region.

Figure 5.2 - Total household position rates by position and sex. North-west area, Year 2002-2040



Source: Authors’ processing on “Aspetti della vita quotidiana” survey data and Households’ projections

To define the LAPs in each projection year, we keep constant the age breakdown for each household position and geographical group and equal to the mean one observed in 2017-2019. Hence, we calculated propensities as follows:

$$LAP_{x,s,i,G,t} = LAP_{x,s,i,G,2017-19} * WP_{s,i,G,t} * WL_{x,s,G,t} \qquad t = 2020, \dots, 2040 \qquad (1)$$

where x=age groups 0-4, ... , 85+, s=sex, i=household position, G=geographical group.

In this formula, $WP_{s,i,G,t}$ is a weight that adjusts $LAP_{x,s,i,G,2017-19}$ on the basis of the future changes in the Total household Position Rates:

$$\mathrm{WP}_{\mathrm{s,i,G,t}} = \frac{\mathrm{TPR}_{\mathrm{s,i,G,t}}}{\mathrm{TPR}_{\mathrm{s,i,G,2017-19}}} \qquad t = 2020, \dots, 2040$$

and $WL_{x,s,G,t}$ is a weight that captures changes in mortality over time:

$$\mathrm{WL}_{\mathrm{x,s,G,t}} = \frac{\mathrm{L}_{\mathrm{x,s,G,2017-19}}}{\mathrm{L}_{\mathrm{x,s,G,t}}} \qquad t = 2020, \dots, 2040$$

At the end of the procedure, the sum of the LAPs by household position in each age group approximates, but does not always equal, the value 100. The problem occurred mainly at the open-aged group (85+) where, because of low absolute frequencies, the impact proved to be not significant. Therefore, some ex-post adjustments were made, which consisted of pro-rating the distributions to the value of 100.

To move from the main geographic groups to regional projections, we allow any single region to keep its own socio-demographic specificity. To that purpose, we introduce a Regional Correction Factor as the ratio between the regional TPR (r) and the macro-regional TPR (G) to which the region belongs, as observed in 2017-2019:

$$\mathbf{RCF_{r,i}} = \frac{\mathbf{TPR_{2017-19,i,r}}}{\mathbf{TPR_{2017-19,i,G}}}$$

where i=household position, r=region, G=macro-region to which region r belongs.

Multiplying the projected LAPs, as from formula (1), by the regional correction factors, we obtain the series of regional LAPs from 2020 to 2040.

As an example, for “lone persons”, the TPR found in Piemonte is 10.79 while in the group 1 is 10.42. The RCF is therefore 1.04. This means that since Piemonte has a higher TPR (more lone persons living there) than the North-west group, it is necessary to make an adjustment by multiplying the projected LAPs by 1.04, increasing their level slightly.

In the fourth step, regional propensities are applied to the projected population living in private households obtained in step 1. This application produces the projected population living in different household positions by sex, age group, and region. Figure 4 in Section 6 shows the age pyramids by household position in 2020 and 2040 at the national level.

In the fifth step, we obtain the number, type and size of projected households. We consider the *Household Representative Rate* (HRR), defined as the probability of a person from a specific group (based on geography, age group, sex, and type of household) being a household reference person. From the population by household position, gender and age, we have:

- each "lone person" represents 1 household (HRR=1);
- a "single parent" acts for 1 household (HRR=1);
- "partners in a couple" constitute 0.5 of a household (HRR=½);
- "children" and "other persons" do not count in the calculation of households (HRR=0).

Moreover, for multi-person households and households with two or more families, the HRR is the ratio between 1 and the average number of people in a multi-person household or in a household with two or more families, as observed from data in 2017-2019. Therefore, the multi-person households were obtained by dividing the number of people living in multi-person households by an average size of this type of household, which has remained broadly stable over time at about 2.1 members. Similarly, the households containing two or more families was derived by dividing the number of persons living in households with two or more families by the average size, which assumed values between 5 and 5.4, depending on the territorial reference group.

Finally, we can calculate household size by dividing the population living in the household by the number of households. With this method, household size can be measured for both total households and households containing at least one family.

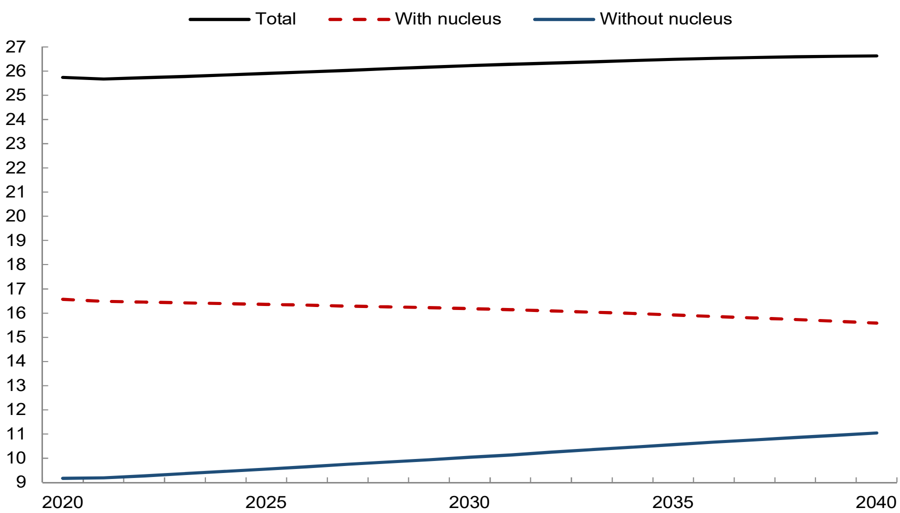
6. Main results

Number of households and size

Projections of the total number of households show an increase of almost one million additional units, following a trend that has already been underway for several years in Italy: from 25.7 million in 2020, it would grow by 3.5 per cent to 26.6 million in 2040.

Such an increase hides a peculiarity of the evolution of families: their fragmentation. Specifically, households without nucleus (non-family households) increase consistently, from 9.2 to 11 million (+20%). Households with at least one nucleus (family households) follow an opposite trend, decreasing from 16.6 to 15.6 million (-6%) (Figure 6.1). This decline is due to the consequences of long-term socio-demographic dynamics, such as the ageing of the population, an increase in marital instability, and low birth rates (Pirani *et al*. 2021). Therefore, an increased life expectancy generates more lone people; the fall in birth rates increases the number of childless people, while the growing marital instability increases the number of people living alone and lone parents.

Figure 6.1 - Projected number of households. Italy, years 2020-2040 (values in millions) (a)



Source: Authors' processing on Households' projections

(a) Data from the survey *Aspetti della vita quotidiana* are disseminated based on a two-year average. Here, however, the data refer to January 1st. As to 2020, this can give rise to differences.

A significant outcome of our predictions is that the mean household size decreases from 2.3 in 2020 to 2.1 in 2040, while the total number of households increases. For households with at least one nucleus, the mean size falls from 3 to 2.8 members over the same period.

Population structure

Past and prospective demographic dynamics in Italy entail a situation in which the number of older people is continuously growing and new generations tend to shrink, both in absolute and relative terms (Istat 2021*a* e 2021*b*). The age structure of the population today shows an imbalance in favour of the older generations and there are no factors that suggest a reversal of this trend. Demographic projections show how unlikely a turnaround in the number of future births can be, even in case of favourable assumptions about fertility (Istat 2021*b*). This is because the prospect of having to deal with a decreasing number of women in childbearing age, on the one hand, and the tendency to postpone parenthood, on the other, seem to be taking on increasing weight.

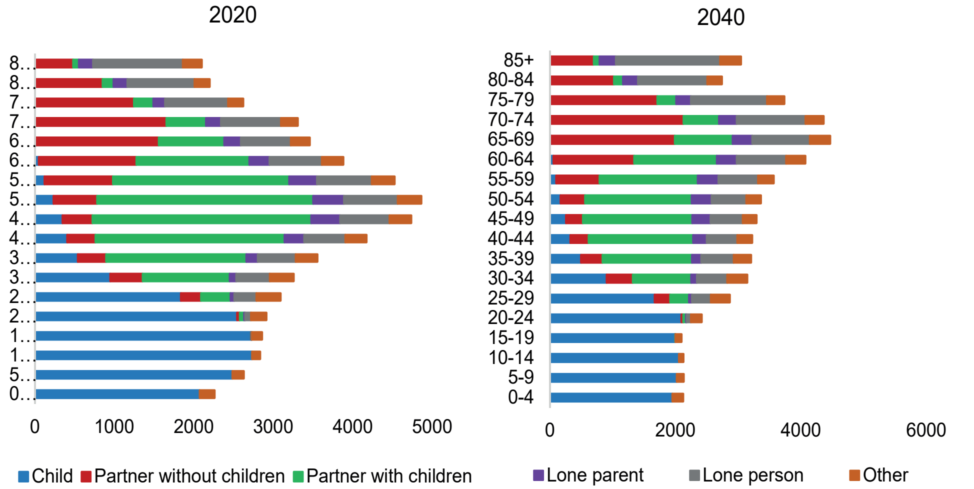
The analysis of the expected population by 2040 according to family role jointly highlights the ageing process and changes in household positions. In particular, it shows a decrease in people living in couples with children, an increase in those without children, and in people living alone, the latter especially if they are older adults (Figure 6.2). The younger age groups are getting smaller, but the “child” family position remains prevalent until the age of 30, reflecting the fact that young people stay longer in their family of origin.

Lone persons

The increase of lone people, real micro-families, is mainly responsible for the absolute growth of the total number of households.

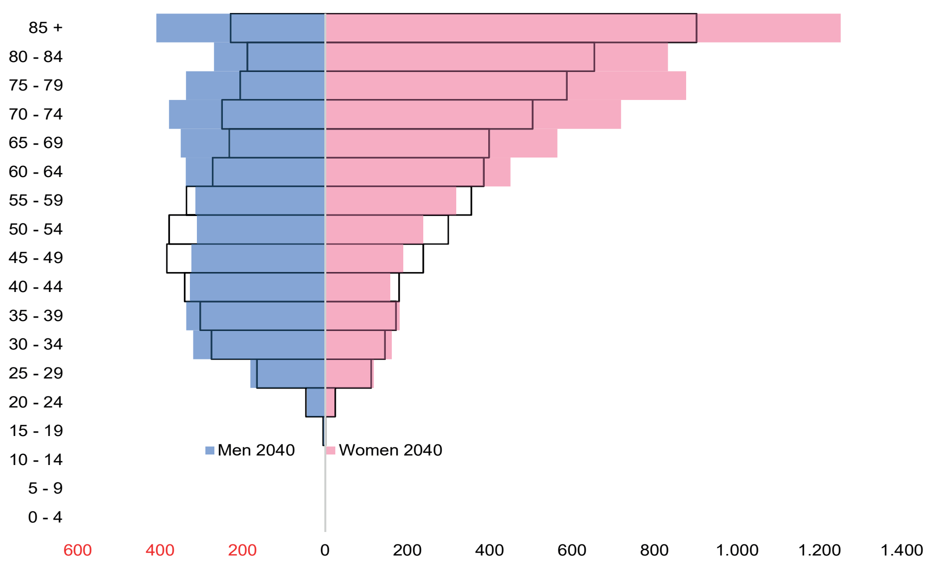
Men living alone are expected from 3.6 million in 2020 to 4.3 million in 2040 (+17%). Women living alone will pass from 5 to 6.1 million, with a 23% growth (Figure 6.3). This growth has a strong social impact, since it is especially in old age that the number of single people increases significantly.

Figure 6.2 - Population by position in the household and five-year group. Italy, years 2020 and 2040 (values in thousands)



Source: Authors' processing on Households' projections

Figure 6.3 - Lone persons by 5-year age group and sex. Italy, years 2020 and 2040 (in thousands)



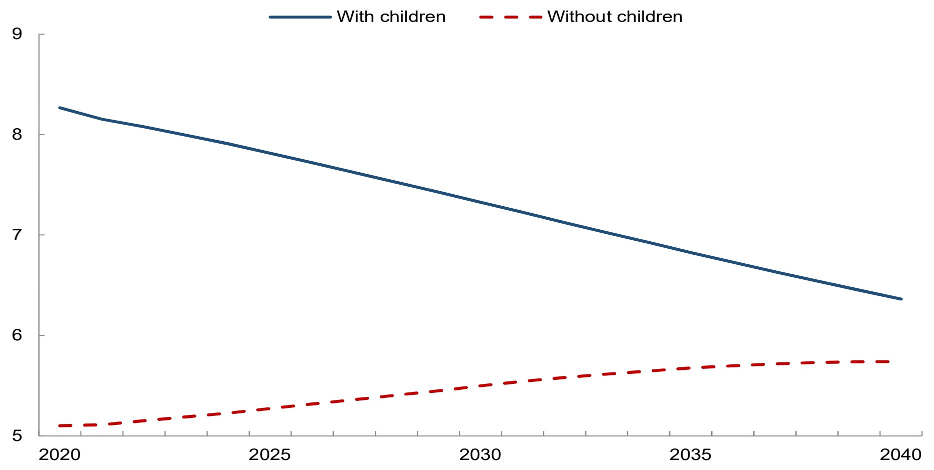
Source: Authors' processing on Households' projections

Among individuals over the age of 64, a spread to 1.2 million lone women and 640,000 lone men are expected. The longer survival of older adults and, among them, of people living alone, could lead to a greater need for care in the future.

Couples with and without children

Due to past and projected fertility levels, couples with children will decrease substantially. Between 2020 and 2040, their number would drop by 23%, from 8.3 million to 6.4 million. At the same time, childless couples are expected to grow slightly, from 5.1 million to 5.7 million, with a 13% increase (Figure 6.4). If these trends continue with the same intensity as predicted up to 2040, especially as regards the decline of couples with children, childless couples could overtake them by 2045.

Figure 6.4 - Couples with and without children. Italy, years 2020-2040 (in millions)



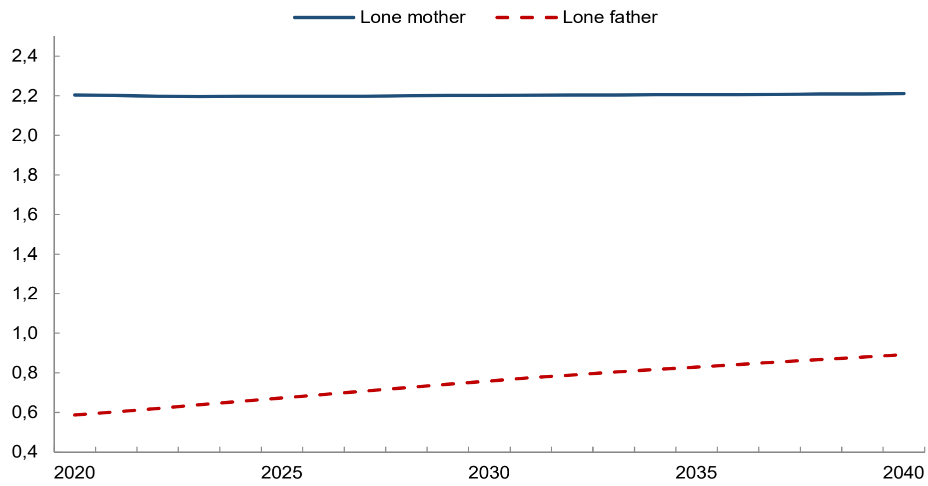
Source: Authors' processing on Households' projections

Single parents

As marital instability becomes increasingly widespread in our country, single parent families will also increase.

In 2020, there were 2.8 million lone parents, more mothers (2.2 million) than fathers (around 600 thousand) and they accounted for 8.6% and 2.3% of total households, respectively. While in the past, after the couple's dissolution, children were generally placed with mothers, the number of fathers as guardians in separation or divorce judgments has increased in recent years[7]. As a result, lone fathers, although still much fewer than lone mothers, would reach about 900,000 (3.4% of total households) by 2040. In that year, single mothers will be numerically unchanged at 2.2 million (8.3% of the total), so that the total number of lone parents is expected to reach 3.1 million (Figure 6.5).

Figure 6.5 - Lone parents by sex. Italy, years 2020-2040 (in millions)



Source: Authors' processing on Households' projections

Geographical differences

At the territorial level, the differences between household types reflect the demographic dynamics and social behaviours typical of the different areas of the country[8].

In the North, in 2020, the share of family households is considerably lower: 62.8% compared to 67.7% in the South (Table 6.1). However, projection

7 This increase is partly due to the 2006 law on joint custody (8 February 2006, no. 54), which provides regulations on the separation of parents and the shared custody of children.

8 The specific territorial classification identified by our model has proven to be particularly effective in modelling the future family formation processes. However, it will not be used for the analysis of the results. In this section, we consider the three large traditional divisions: North, Centre and South, in line with the usual Italian data release, thus enabling a comparison of time series. The main difference between the two classifications concerns the Centre, whose regions belong to Groups 2, 3 and 5. The North roughly corresponds to Groups 1 and 2, while the South refers mostly to Groups 4 and 5.

results show a tendency for convergence between the two areas. In the South, in fact, a more consistent change in this type of household is expected, since in 2040 they could constitute 61% of the total households (a reduction of about 7 percentage points). In the North, in contrast, non-family households will see a smaller reduction, reaching 57.5% of total households in 2040.

Table 6.1 - Households by type and geographic area. Years 2020, 2030, 2040 (percentage values)

Household type	North			Centre			South			Italy		
	2020	2030	2040	2020	2030	2040	2020	2030	2040	2020	2030	2040
Lone man	14.7	15.8	17.0	15.0	15.6	16.7	12.5	13.0	13.9	14.1	14.8	16.0
Lone woman	20.3	21.5	22.9	19.9	21.7	23.4	17.3	19.6	22.4	19.2	20.9	22.8
Childless couple	21.8	22.7	23.1	18.2	19.2	19.6	17.9	19.5	20.4	19.8	21.0	21.6
Couple with children	30.0	26.2	22.8	29.7	25.7	21.9	36.9	32.0	26.9	32.1	27.9	23.9
Lone father	2.2	2.8	3.3	2.6	3.3	3.7	2.2	2.8	3.2	2.3	2.9	3.4
Lone mother	7.5	7.3	7.1	10.3	10.1	10.3	9.0	8.9	8.9	8.6	8.4	8.3
Other type	3.5	3.7	3.8	4.3	4.5	4.6	4.3	4.3	4.3	3.9	4.0	4.1
Households with nuclei	62.8	60.3	57.5	62.9	60.3	57.2	67.7	64.8	61.0	64.4	61.7	58.5
Households without nuclei	37.2	39.7	42.5	37.1	39.7	42.8	32.3	35.2	39.0	35.6	38.3	41.5

Source: Authors' processing on Households' projections

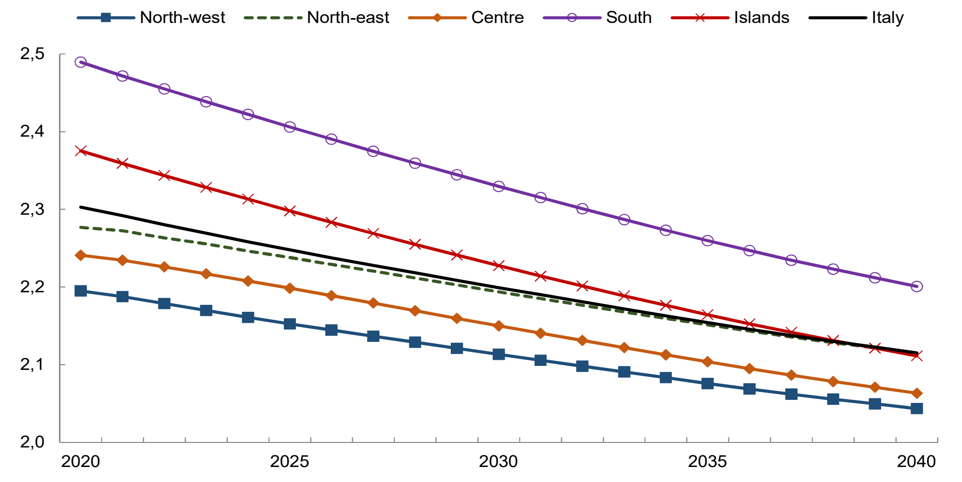
In the short term, the ageing process will be particularly intense in the South. Although this area still has a younger structural profile, its mean age will rise from 44.6 years in 2020 to 50 years in 2040, surpassing even the North, where it will reach 49.2 years (from an initial level of 46.3).

Spatial and gender gaps in survival will affect the growth of people living alone: women will increase from 20.3 to 22.9% (+2.6%) in the North and from 17.3 to 22.4 (+5.1%) in the South, leading to a convergence between the two areas. The proportion of men living alone, meanwhile, will remain at a much lower level in the South than in the North (in 2040, 13.9% vs. 17%). This result is probably due to men's greater propensity to contract second marriages or to live with other relatives in the South rather than in the North (Rettaroli 1997; Angeli *et al*. 2003; Meggiolaro *et al*. 2008; Gałęzewska et al. 2017). As already mentioned, couples with children are the household type that will undergo the greatest change over the next 20 years, falling from 32.1% to 23.9% of total households. In the South, this decrease would be more pronounced, with a drop of 10 percentage points (from 36.9% to 26.9%), due to the strong decrease in fertility rates in these regions.

Childless couples will continue to be more widespread in the North (23.1%), although with a smaller increase. In fact, the most significant change is expected in the South, where, despite a less widespread starting situation, couples without children would increase by about 3 percentage points in the next twenty years (from 17.9% to 20.4%).

All these changes in family structures will have an impact on the households' average size, which will continue to fall according to territorial specificities. The North and the Centre, with very similar current values and future trajectories, will reach an average number of members just below the national value. The South, thanks to historically higher fertility rates, has always had larger families than the North. Today, with declining reproductive levels even in the South, this primacy (2.5 members) tends to become less clear. In the future, the expectation is for a further decline to 2.2 members (Figure 6.6).

Figure 6.6 - Mean household size by geographic area. Years 2020-2040



Source: Authors' processing on Households' projections

7. Conclusions and perspectives

Household projections represent fundamental and priority information for policy planning. Despite the vast utility of knowing the family structure of the future population, official household forecasts are still not widespread. In 2020, the Italian National Statistical Institute started a project with the aim of releasing official household projections integrated with the demographic projections on an annual basis.

In this paper, we described the methodology implemented for projecting the number and type of households in our country over a 20-year time horizon. A static method based on the propensity-rate model was used. The method consists of a few simple steps. First, the projected population is broken down into persons living in private and institutional households; then, propensity rates for the entire projection period are calculated and predicted to obtain the projected population by living arrangements. Finally, predictions on the number of future households, their average size and composition are derived.

The main strengths of this method are: it easily links to existing population projections; data inputs are not burdensome and allow projections to be annually updated; it provides quite detailed results. Starting from the approach developed by the Australian Bureau of Statistics – ABS (2019), we supplemented it with some methodological refinements. These include the introduction of a synthetic indicator, the Total household position rate, regardless of age, and the territorial top-down strategy. The former made it possible to better translate the assumptions about family behaviour and to keep control on the predicted trends of the various household positions. The latter permitted to derive robust and consistent results at the regional level, considering that the release of territorial information is essential for a country like Italy, where geography is itself an interpretative key more than a classification variable.

However, we are still working on some open issues. First, the two-sex problem, that arises when males and females are modelled separately, so that the predicted number of male partners is not equal to the number of female partners. It has proven to be a rather difficult problem, because it has both conceptual and methodological aspects and there is no simple way to bring empirical data to bear on it. Although there are several interesting proposals

(Schoen 1988), we had to overlook them, as it was essential to maintain numerical consistency between the population by household position and the total predicted population by sex and age. Furthermore, we intend to provide measures of the uncertainty of the estimates, which are important for better understanding phenomena and making appropriate decisions.

The results show an expected increase by one million in the number of households together with a decrease in the mean household size that would drop from 2.3 to 2.1 persons in a 20-year time horizon. We find a reduction in the share of couples with children and a growing importance of couples without children and of persons living alone. By 2040, only one in four families will consist of a couple with children, while more than one in five will be childless. In addition, more than 10 million people will live alone in 2040, from an initial value of 8.6 million in 2020. A large proportion of these lone people will be older and this may lead to a greater need for care in the future. However, more old people in the population may also have positive implications: the increase in the number of total years lived, with many of them in good health conditions, could enable these persons to play an active role in the society. For example, as is already the case today and more likely in the future, by supporting their children's families in taking care for their grandchildren and providing them with economic support, participating in the economic cycle not only as consumers of welfare services, but also as capital investors (Istat 2020; Quattrociocchi *et al.* 2021).

Although the results are widely expected, also because they are in line with the overall picture that emerges from the general demographic trends, we believe that our analysis poses a series of useful indications for the planning of policies. In particular, we highlight the need to pay greater attention in the coming years to young families with children and to those that include very old people. Their needs may turn out to be crucial not only for themselves, but also for accompanying a sustainable development of the country system as a whole.

For the future, the outcome of the model and the feedback received from users represent a further incentive for us to refine some methodological issues that would allow better investigating the structure of the Italian household. Among them, for example, we aim at getting more information on couples with children and single parents, by identifying those with children by age group (or below a given age threshold). On the other hand, looking at the issue of the progressive ageing of the population and to the growing fragility of very old people, we are interested in better understanding the family context of older adults, whether they live in couple or alone, or live with the present support of their children.

References

Angeli, A., e A. De Rose. 2003. “Donne e uomini dopo lo scioglimento della prima unione”. In Pinnelli, A., F. Racioppi, e R. Rettaroli (*a cura di*). *Genere e Demografia*: 367-392. Bologna, Italia: il Mulino.

Australian Bureau of Statistics - ABS. 2019. *Household and family projections for Australia, states, territories and capital cities 2016-2041*. Canberra, Australia: ABS. https://www.abs.gov.au/statistics/people/population/household-and-family-projections-australia/2016-2041#:~:text=Data%20Explorer%20datasets-,Key%20statistics,the%20most%20common%20living%20arrangement.

Bagatta, G. (*a cura di*). 2006. “Il sistema di indagini sociali multiscopo. Contenuti e metodologia delle indagini”. *Metodi e Norme*, N. 31. Roma, Italia: Istat. https://www.istat.it/wp-content/uploads/2014/06/met_norme_06_31_il_sistema_di_indagini_multiscopo.pdf.

Bell, M., J. Cooper, and M. Les. 1995. *Household and family forecasting models: A review*. Canberra, Australia: Commonwealth Department of Housing and Regional Development.

Blangiardo, G., E. Barbiano di Belgiojoso, e P. Bonomi. 2012. “Le previsioni demografiche delle famiglie”. In Donati, P. (*a cura di*). *La famiglia in Italia. Sfide sociali e innovazioni nei servizi. Volume I: Aspetti demografici, sociali e legislativi*: 91-123. Roma, Italia: Carocci.

Box, G.E.P., G.M. Jenkins, G.C. Reinsel, and G.M. Ljung. 2015. *Time Series Analysis: Forecasting and Control. 5th Edition*. Hoboken, NJ, U.S.: John Wiley and Sons.

Cannari, L., and S. Nicoletti Altimari. 1998. “A Dynamic Micro Simulation Model of the Italian Households’ Sector”. In Baldacci, E., e F. Peracchi (*a cura di*). “Le previsioni della spesa per pensioni. Metodologie a confronto”: 103-136. Atti del Workshop, Roma 2 dicembre 1997. *Annali di Statistica*, Anno 127, Serie X - vol. 16. Roma, Italia: Istat. https://ebiblio.istat.it/digibib/Annali/TO00003841_Serie10Vol16Ed1998.pdf.

Castagnaro, C., e E. Meli (*a cura di*). 2022. “Famiglie, reti familiari, percorsi lavorativi e di vita”. *Letture Statistiche – Temi*. Roma, Italia: Istat. https://www.istat.it/it/archivio/275924.

de Beer, J.A.A., and M. Alders. 1999. "Probabilistic population and household forecasts for the Netherlands". *Working Paper*, N. 45. Joint ECE-Eurostat *Work Session on Demographic Projections*, Perugia, Italy, 3-7 May 1999. https://unece.org/fileadmin/DAM/stats/documents/ece/ces/1999/05/projections/45.e.pdf.

Décarie, Y., M. Boissonneault, and J. Légaré. 2012. "An Inventory of Canadian Microsimulation Models". *Social and Economic Dimensions of an Aging Population Research Papers*, N. 298. Hamilton, Canada: McMaster University.

Gill, R., and N. Keilman. 1990. "On the estimation of multidimensional demographic models with population registration data". *Mathematical Population Studies*, Volume 2, N. 2: 119-143.

Gałęzewska, P.B., B. Perelli-Harris, and A. Berrington. 2017. "Cross-national differences in women's repartnering behaviour in Europe: the role of individual demographic characteristics". *Demographic Research*, Volume 37, N. 8: 189-228.

Hantrais, L., D. Philipov, F.C. Billari. 2006. "Policy implications of Changing Family Formation. Study Prepared for the European Population Conference 2005". *Population Studies*, N. 49, Strasbourg, France: Council of Europe Publishing.

Hays, J.C. 2002. "Living arrangements and health status in later life: a review of recent literature". *Public Health Nursing*, Volume 19, N. 2: 136-151.

Istituto Nazionale di Statistica - Istat. 2021*a*. "Indicatori demografici. Anno 2020". *Statistiche Report*. Roma, Italia: Istat. https://www.istat.it/it/files/2021/05/REPORT_INDICATORI-DEMOGRAFICI-2020.pdf.

Istituto Nazionale di Statistica - Istat. 2021*b*. "Previsioni della popolazione residente e delle famiglie. Base 1/1/2020". *Statistiche Report*. Roma, Italia: Istat. https://www.istat.it/wp-content/uploads/2021/11/REPORT-PREVISIONI-DEMOGRAFICHE.pdf.

Istituto Nazionale di Statistica - Istat. 2021*c*. "Nota tecnica sulla produzione dei dati del Censimento Permanente: la popolazione residente per genere, età, cittadinanza e grado di istruzione al 31.12.2020". *Censimenti Permanenti. L'Italia giorno dopo giorno*. Roma, Italia: Istat. https://www.istat.it/wp-content/uploads/2021/12/NOTA-TECNICA-CENSIMENTO-POPOLAZIONE_2020.pdf.

Istituto Nazionale di Statistica - Istat. 2020. "Invecchiamento attivo e condizioni di vita degli anziani in Italia". *Letture Statistiche - Temi*. Roma, Italia: Istat. https://www.istat.it/produzione-editoriale/invecchiamento-attivo-e-condizioni-di-vita-degli-anziani-in-italia/.

Keilman, N. 2019. "Family Projection Methods: A Review". In Schoen, R. (*Ed.*). *Analytical Family Demography*: 277-301. Cham, Switzerland: Springer, *Series on Demographic Methods and Population Analysis*.

Keilman, N. 1985. "Nuptiality models and the two-sex problem in national population forecasts". *European Journal of Population/Revue Européenne de Démographie*, Volume 1, N. 2-3: 207-235.

Keilman, N., A. Kuijsten, and A. Vossen (*Eds.*). 1988. *Modelling household formation and dissolution*. Oxford, UK: Oxford University Press (Clarendon Press).

Kono, S. 1987. "The Headship Rate Method for Projecting Households". In Bongaarts, J., T.K. Burch, and K.W. Wachter. *Family Demography: Methods and Their Applications (International Studies in Demography)*: 287-308. Oxford, UK: Oxford University Press (Clarendon Press).

Lavit, C., Y. Escoufier, R. Sabatier, and P. Traissac. 1994. "The ACT (STATIS) method". *Computational Statistics & Data Analysis*. Volume 18, N. 1: 97-119.

Linke, W. 1988. "The headship rate approach in modelling households: The case of the Federal Republic of Germany". In Keilman, N., A. Kuijsten, and A. Vossen (*Eds.*). *Modelling household formation and dissolution*: 108-122. Oxford, UK: Oxford University Press (Clarendon Press).

McDonald, P., R. Kippen, and J. Temple. 2006. “Net transition probabilities: an approach to subnational level projections of households and housing demand based on census data”. *Population, Space and Place*, Volume 12, N. 6: 479-495.

Meggiolaro, S., and F. Ongaro. 2008. “Repartnering after marital dissolution: Does context play a role?”. *Demographic Research*, Volume 19, Article 57:1913-1934. https://www.demographic-research.org/volumes/vol19/57/19-57.pdf.

Murphy, M. 1991. “Household Modelling and Forecasting-Dynamic Approaches with Use of Linked Census Data”. *Environment and Planning A: Economy and Space*, Volume 23, N. 6: 885-902.

O’Neill, B.C., and B.S. Chen. 2002. “Demographic Determinants of Household Energy Use in the United States”. *Population and Development Review*, Volume 28, Supplement *Population and Environment: Methods of Analysis*: 53-88.

Paciorek, A.D. 2013. “The Long and the Short of Household Formation”. *FEDS Working Paper*, N. 2013-26. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2269052#.

Pirani, E., G. Guetto, e F. Rinesi. 2021. “Le famiglie”. In Billari, F.C., e C. Tomassini (*a cura di*). *Rapporto sulla Popolazione. L’Italia e le sfide della demografia*. Bologna, Italia: il Mulino.

Quattrociocchi, L., M. Tibaldi, M. Marsili, L. Fenga, and M. Caputi. 2021. “Active Ageing and Living Condition of Older Persons Across Italian Regions”. *Journal of Population Ageing*, Volume 14, N. 1: 91-136.

Rettaroli, R. 1997. "Le seconde nozze". In Barbagli, M., e C. Saraceno (*a cura di*). *Lo stato delle famiglie in Italia*: 257-271. Bologna, Italia: il Mulino.

Schoen, R. 1988. *Modeling Multigroup Population*. New York, NY, U.S.: Springer.

Stevens, B., P.-A. Schieb. 2012. “The future of families to 2030: an overview of projections, policy challenges and policy options”. In Organisation for Economic Co-operation and Development - OECD. *The Future of Families to 2030*: 15-55. Paris, France: OECD Publishing. https://doi.org/10.1787/9789264168367-en.

Sullivan, D.F. 1971. "A single index of mortality and morbidity". *HSMHA Health Reports*, Volume 86, N. 4: 347-354

United Nations. 1973. "Manuals on methods of estimating population. Manual VII - Methods of Projecting Households and Families". *Population Studies*, N. 54. Geneva, Switzerland: United Nations Publication.

van Imhoff, E., and N. Keilman. 1991. *LIPRO 2.0: An application of a dynamic demographic projection model to household structure in the Netherlands*. Amsterdam, The Netherlands: A. A. Balkema.

Willekens, F.J.C., J.A.A. de Beer, and N.L. van der Gaag. 2007. *MicMac: bridging the micro-macro gap in population forecasting: report on input data requirements of MIC: work package 2, micro-simulation model (Deliverable D.9)*. The Hague, The Netherlands: Netherlands Interdisciplinary Demographic Institute - NIDI.

Wilson, T. 2013. "The sequential propensity household projection model". *Demographic Research*, Volume 28, Article 24: 681-712. https://www.demographic-research.org/volumes/vol28/24/28-24.pdf.

Zeng, Y., K.C. Land, Z. Wang, D. Gu. 2013. "Household and Living Arrangement Projections at the Subnational Level: An Extended Cohort-Component Approach". *Demography*, Volume 50, N. 3: 827-852.

An analysis of the demand for sub-municipal data from the Population and Housing Census

Giancarlo Carbonetti[1], Antonella Ciccarese[1], Roberta Roncati[1]

Abstract

Starting in 2018, the Italian National Institute of Statistics - Istat launched the new Permanent Census of Population and Housing, thus moving from the traditional census to a combined census based on the integration of administrative and sample data.
This change of strategy has an important impact on the data offered: due to the methodological complexity required by the new census approach, it is indeed not possible to guarantee data at the highest territorial detail in continuity with past censuses.
As requests for data from the 2011 Census often present a high degree of classificatory and spatial detail, it seemed appropriate to assess whether the dissemination plan for census data can meet the expectations of specialised users of sub-municipal data processing, according to Principle n. 11 of the European Statistics Code of Practice.
To this end, a fact-finding survey was conducted by interviewing advanced users who had previously requested the 2011 Census data at the sub-municipal level, intending to assess satisfaction with the quality of the data received and the characteristics of future requests referring to the new Permanent Census.
The analysis shows that NSIs have to concentrate great efforts, both in data collection and in defining statistical methodologies, so that the provision of sub-municipal data related to censuses is relevant to user needs.

Keywords: Advanced users, custom data processing, data dissemination, data quality, population census, relevance, spatial data, survey, users' needs.

DOI: https://doi.org/10.1481/ISTATRIVISTASTATISTICAUFFICIALE_1-2-3.2023.02.

---

1 Giancarlo Carbonetti (carbonet@istat.it); Antonella Ciccarese (antonella.ciccarese@istat.it); Roberta Roncati (roncati@istat.it), Italian National Institute of Statistics – Istat.

The views and opinions expressed are those of the authors and do not necessarily reflect the official policy or position of the Italian National Institute of Statistics - Istat.

The authors would like to thank the anonymous reviewers for their comments and suggestions, which enhanced the quality of this article.

1. Introduction

The demand for quantitative information at a high level of territorial detail has been satisfied over time by the results of the decennial Census, the only statistical survey capable of producing data at the sub-municipal level. In response to this demand, up to the 2011 Census of Population and Housing, the Italian National Institute of Statistics – Istat disseminated data on more than 400,000 enumeration areas and satisfied requests for specific processing received through its Contact Centre service (Istat 2022; Istat 2021).

In Italy, since 2018 Istat has abandoned the traditional census to start the new Permanent Population and Housing Census[2], a "combined" census based on the integration of administrative and sample data (Falorsi 2017). The new strategy has a strong impact on data supply due to the methodological and computational complexity required by the new census approach. Indeed, the process of producing the results of the new Permanent Census will not be able to provide data at the highest territorial detail in continuity with past censuses.

Unlike past censuses, for which users could request customised processing of sub-census data, with the new census a set of results will be defined, growing over time, that users use to get the information they are interested in. It is therefore important that this set is as close as possible to users' expectations.

Indeed, the European Statistics Code of Practice (last revised in 2017) sets out the principles to be applied to ensure and strengthen both trust and quality in the European Statistical System (ESS and Eurostat 2017). Among the principles of the Code - largely inspired by the Fundamental Principles of Official Statistics adopted by the United Nations General Assembly in 2014 - Article 11 emphasises the relevance of statistics and the need to meet the needs of users:

2 Since 2018, Istat has been conducting in Italy the Permanent Census of Population and Housing. The traditional decennial census essentially based on collecting data from people, has been replaced by a census based on a system of registers supported by sample surveys. Every year counts at municipal level are disseminated according to the Basic Register of Individuals (BRI), the Basic Register of Places (BRP) and a Population Coverage Survey (PCS). BRI contains information on some demographic variables such as gender, place and date of birth, citizenship, place of residence, derived by administrative data. BRP contains addresses, Enumeration Areas (EAs) and if possible, geographical coordinates. All other census variables not included in the registers are collected with the traditional census questionnaire each year on household samples on representative sets of municipalities. From the integration of the data in the registers and the data collected on the sample households, census results are produced every year for different information details at the various territorial levels. This change required the adoption of new methodological and IT architectures with the aim of providing accurate, timeliness and consistent figures for the users.

- Procedures are in place to consult users, monitor the relevance and value of existing statistics in meeting their needs, and consider and anticipate their emerging needs and priorities. Innovation is pursued to continuously improve statistical output.
- Priority needs are being met and reflected in the work programme.
- User satisfaction is monitored on a regular basis and is systematically followed up.

Within this framework, Istat conducted the study presented here to gain a clear view of what the expectations of stakeholders are concerning high spatial detail data. To this end, advanced users who had requested the 2011 Census data at the sub-municipal level were interviewed to assess satisfaction with the quality of the data received (Department of Public Service/*Dipartimento della Funzione Pubblica* 2004) and the characteristics of future requests for the new Permanent Census.

At the same time, research was conducted on NSIs' experiences in other countries regarding user surveys on the level of satisfaction with or requirements for census data with a high spatial detail. Unfortunately, no user analysis comparable to the one described in this article was found.

2. Satisfaction surveys on statistical products and services conducted by NSIs

In many countries, National Statistical Institutes (NSIs) conduct user satisfaction surveys of products and services on a regular or occasional basis.

Attention to users and their needs has a long tradition and has consolidated over time to the point of changing the relationship between institutions and users themselves. The latter are no longer seen as mere recipients of the actions of the statistical institute, but as stakeholders from whom to learn to improve the quality of products and services. Their opinions can help identify the factors where the largest gap is registered between what the institution achieves and what users need or expect to receive.

The most important aspects that are measured through user ratings are timeliness, relevance of the data offered and, in general, the quality of the statistics produced, which must meet the information needs of society. Indeed, user feedback could be integrated into the data production processes and planning of official statistics. Users should be placed at the centre of statistical production: their needs should be understood, their opinions sought and considered, and their use of statistics supported.

The NSIs of certain countries[3] - *e.g.* Albania, Bulgaria, Canada, Estonia, Greece, Ireland, and Slovenia - collect information from users in an anonymous form, according to a continuous flow when accessing the institutional website or dissemination web platform, after consultation or downloading of statistical data. Periodically, this information is processed for internal use and in some cases presented in technical reports available on the web.

In other countries - *e.g.* Cyprus, Ghana, Rwanda, Serbia, Spain, Tanzania, and the United Kingdom - specific surveys are occasionally conducted on samples of advanced users from different economic sectors, including central or local government, academia, business, the voluntary sector, precisely to gather feedback from users with high statistical and IT skills who can use the data for policy, study and planning purposes.

3 In the references at the bottom of the article are the links to the websites of each country cited.

These surveys mostly consider the entire statistical output and no specific attention has ever been paid to population and housing census results or data with a high spatial detail. It was therefore not possible to observe any analysis experience on advanced users similar to that reported in this work, which is the first on a specialised user that requires and uses census data with high information and spatial content. The return of the information from this survey, even if it refers to a limited sample size, is crucial for guiding future choices in terms of the production process and planning the release of official statistics.

3. Knowledge objectives

The cornerstones of the “Population Census - User Satisfaction Survey” (Carbonetti *et al*. 2022) were built on precise questions:

- Did the data provided fully satisfy the demand?
- Did the data provided make it possible to realise the objectives of study, research, analysis, and planning that the users had set themselves?
- Did the data provided stimulate new and different projects not initially envisaged?
- Did the level of detail and quality of the data provided, sometimes limited by privacy constraints, really meet the users’ needs?
- Were the data also used through the use of GIS?
- Has the data received been integrated with other sources available to the user?
- Will users in the future need to request new supplies of similar or more detailed data from a spatial or classification point of view?

This last question is linked to the opportunities that the new Permanent Population and Housing Census will be able to offer. There is a strong expectation on the part of external users of highly detailed and timely spatial information that can be obtained every year thanks to the new census operation put in place.

The Istat Contact Centre monitors and testifies to the need for “continuity” expressed by specialist and loyal user segments that, thanks to the population census, benefit from spatial information down to the smallest detail, in historical series, filling documentary “gaps” that other surveys, due to their methodology, cannot cover.

The experience gained in this context stimulated the idea of a survey of all users who purchased customised elaborations on the 15th Census of Population and Housing from 2015 to 2021 with the following peculiarities:

- high spatial detail (sub-municipal: by census area or by enumeration area);
- high information detail (often involving crossings of census variables not included in the Italian Dissemination Plan);
- absence on Istat’s dissemination platform.

The survey was therefore conducted on a particular user group, interested in data with a high information content in both spatial and classificatory terms. The requests themselves imply particular characteristics of the users in terms of:

- finality;
- specialisation in data processing;
- attention to quality;
- propensity to demand for new data.

Therefore, given the peculiarities of the data requested and the characteristics of the users who requested them, a need for knowledge emerged to study the specific user segment and try to assess:

- relevance and satisfaction of the data provided;
- achievement of goals and any other objectives;
- possibility of future requests for similar data or different details.

Trusting in a significant response rate, the return of the survey aims to "identify" the information needs of census data, also in the perspective provided by the new Permanent Census of Population and Housing in Italy, which renews its offer from year to year, to assess how far Istat will be able to satisfy the need for information with a strong spatial detail.

4. Handling of customised processing requests

Since the implementation of the 2011 Census of Population and Housing planned the adoption of a sample strategy based on short/long forms (Borrelli *et al.* 2011; Carbonetti and Fortini 2008; Carbonetti *et al.* 2008) for the determination of the census results referring to the different spatial and classificatory levels it was decided to use “constrained weighting estimators” that required the calculation of specific “carry-over weights to the universe”, associated with the statistical units, and calculated separately for individuals, households, and dwellings.

The methodology put into production allowed for the consistent determination of all the census crossings scheduled by the Italian Diffusion Plan, concerning the calibration constraints, the different classification hierarchies (data breakdowns) and different spatial levels (Borrelli *et al.* 2012; Carbonetti 2009; Carbonetti and Verrascina 2010).

The weak point of the estimation methodology adopted lies in a certain “rigidity” about the set of census crossings and spatial levels predefined by the Italian Diffusion Plan. Each census crossing not included in the Italian Diffusion Plan required an evaluation of compatibility with the System of Constraints adopted for the calculation of the carry-over weights to the universe; in the event of non-compatibility, one had to proceed to the calculation of appropriate *ad hoc* weights, a different weighting system based on a specific system of constraints. This is what happened with customised processing requests: requests for the supply of data referring to census crossing are often not foreseen in the publication plan and refer to municipal or sub-municipal domains.

To meet these requests, the Census Department developed in 2015 a generalised methodology based on the reweighting procedure, which allows, starting from the calibration weights already prepared, to recalculate a new system of weights, depending on the request taken in charge, useful for producing the required crossings as consistent as possible, at the various levels of classificatory and/or spatial aggregation, with what has already been published (small deviations are due solely to rounding).

Thus, already in the second half of 2015, the Institute was able to support customised processing requests received through its Contact Centre. For each request, a feasibility analysis was carried out in computational terms and a cost estimate was drawn up in terms of working hours to carry out the requested processing and delivery time. When the user, informed by the Istat Contact Centre of the outcome of the assessment, accepted the cost estimate, the Census Department proceeded with the necessary processing to complete the requested delivery within the stipulated time.

In some situations, methodological and thematic experts provided crucial technical and scientific support to assess the relevance of the request concerning the stated objectives and to inform users about the risks of too much detail in terms of data fragmentation (the main risk was that of privacy violation). This activity sometimes led to redesigning the initial requests together with the user.

5. Survey design

5.1 Questionnaire

The questionnaire was designed to collect useful data to meet the information needs set out in paragraph 3. It includes:

1. two closed-ended questions on the user's profile and the reason for the data request;
2. a box for the description of the field or topic for which the data were requested;
3. six closed-ended questions on the use of the data, their level of quality and the possibility of making similar requests about the results of the new Permanent Population and Housing Census;
4. a question on the overall quality of the service received from Istat;
5. a final box in which to report critical issues or provide suggestions.

The individual questions and, in the case of closed-ended questions, the possible answer options are listed in the Annex at the bottom of this article.

5.2 User list

As mentioned above (see paragraph 3), the field of observation for the survey was restricted to users of the Istat Contact Centre who, between 2015 and early 2021, purchased customised processing of 2011 Census of Population and Housing data, referring to sub-municipal domains (census areas; enumeration areas) for census crossings involving at least one estimated variable[4].

4 A sample strategy based on short/long form was adopted for the 2011 Census of Population and Housing. In provincial capitals and municipalities with more than 20,000 inhabitants (486 in total), the census design provided for the long form questionnaire to be administered only to suitably selected samples of households, while the non-sampled households were surveyed using the short form questionnaire. In all other municipalities, the survey was conducted using only the long questionnaire. To produce the final results, for the sampled municipalities the data referring to the variables present in the long questionnaire were estimated, while those relating to the variables present in both models were derived from a counting operation; in the non-sampled municipalities, on the other hand, all data were the result of a counting operation.

The survey was aimed precisely at users specialised in advanced data processing - who are the main stakeholders of census data - to study, based on an analysis of their needs, the relevance of the data that will be produced and offered with the Permanent Census. The need to focus on such a small set of users led to the consideration of a limited but highly representative (reasoned) sample of users for the purposes of the study.

Therefore, regarding the period defined above and for the set of users specified, the number of requests for customised paid processing received by the Contact Centre was 77: for 14 of these the cost estimate was not accepted and, consequently, the data were not provided; for 19 others, instead, these were requests with an accepted estimate but referring to users who had already acquired data on the occasion of previous requests. Thus, excluding users who did not receive data due to non-acceptance of the cost estimate and cases of subsequent requests referring to the same user, the final number of potentially eligible users[5] for the survey is 44. Table 5.1 shows the different cases just described also concerning the year in which the request reached the Contact Centre (in the case of several requests made by the same user, reference is made to the year of the last request).

Table 5.1 - Number of Contact requests received, not executed, multiple. Number of users potentially eligible for the survey, by year of Contact request (absolute values)

Requests for processing of sub-municipal data estimates from the 2011 Census	Year of request to Contact Centre							
	2015	2016	2017	2018	2019	2020	2021	**Total**
No. of requests received by the Istat Contact Centre	9	17	12	15	12	10	2	**77**
No. of requests not supplied due to non-acceptance of the cost estimate	2	5	3	2	-	1	1	**14**
No. of requests following a request already provided for the same user	2	7	3	3	3	1	-	**19**
No. of users who made one or more requests with reference to the year of the one/last request (users eligible for the survey)	5	5	6	10	9	8	1	**44**

Source: Istat - Directorate for Communication, Information and Services to Citizens and Users - Management and Dissemination Service

5 The user is only declared definitively eligible if there are valid contact details to reach him/her and involve him/her in the survey.

5.3 Survey technique

For the data collection phase, the CAWI (Computer Assisted Web Interviewing) methodology was adopted, a data collection technique based on the completion of a web-based questionnaire provided via a link to a website. With the support of the Institute's IT services[6], the questionnaire described in the Annex was prepared in electronic format on the LimeSurvey application[7]. Users potentially eligible to participate in the survey were sent an information letter (see paragraph 6) via e-mail in which, after explaining the purpose of the survey, a link was provided to access the LimeSurvey service where they could fill out the questionnaire themselves.

6 The electronic questionnaire was carried out using the LimeSurvey application with the contribution of Andrea Nunnari (Istat, Directorate for Information Technology).

7 LimeSurvey is an application that allows the creation of online questionnaires and statistical surveys.

6 Stages of the survey

The first operation conducted was to find all the contact information (email address; telephone number) of the advanced users assumed to be eligible for the survey (44). Subsequently, the following "information letter" was sent out in which the user was informed of the reason for the survey, even years after they had requested the data, and invited to answer the questions in the questionnaire accessible through the web application link given at the end of the letter.

> *Dear user,*
>
> *following the dissemination of the results of the 15th Census of Population and Housing in 2011, the National Statistical Institute provided data referred to thematic crossings and spatial domains that were not included in the Italian Dissemination Plan and therefore not available on the Institute's dissemination platforms.*
>
> *Since you have requested customised processing of 2011 Census data through the Istat Contact Centre, we would like to ask for your willingness to participate in a survey whose sole objective is to assess the degree of satisfaction and the level of use of the data provided in order to accurately plan the future production of census data with a high level of informative detail.*
>
> *Thanking you in advance for your attention, we ask you to answer the following short questionnaire.*
>
> **Survey access link**

Following the sending of this letter, the contact e-mail of 7 users was found to be incorrect or non-existent. After the patient work to retrieve the new contact information, almost all critical cases were resolved except for two users who were not found. Consequently, the final target of the survey was set at 42 eligible users (Table 6.1).

The survey started in the last week of May 2021 and lasted approximately one and a half months. After two reminders to non-respondents, the survey ended in the second week of July. After a strong start in the first week (response rate of 47.6%), following the first reminder sent by e-mail to all users (which took place 10 days after the start of the survey), the response rate increased to 64.3%. After the second reminder, sent only to non-respondents exactly 4 weeks after the start of the survey, the response rate increased by approximately 12 percentage points to a final figure of 76.2% (32 out of 42 respondents).

Of the 42 users invited to participate, 32 responded positively (76.2% response rate). Ten users (23.8%) did not participate. Taking the year of the last request to the contact centre as a reference, no relationship is observed between the number of years since the last request and the response rate. A negative effect was expected from the memory of the data request, which might have led to a lower propensity to participate in the survey in the case of very old requests. Instead, it can be seen that (Table 6.1) for 2017, 2020 and 2021 requests, participation was 100%, while lower response rates were recorded for 2016 (50%) and 2019 (55.6%).

This therefore suggests that, also in the future, in the case of surveys on the satisfaction of services provided by Istat to external users, the time factor does not seem to affect the response rate, so it is always possible to retrieve information from users who came into contact with Istat several years earlier.

Table 6.1 - Number of initial users, unreachable users, users contacted for the survey, and responding users, by year of request to the Istat's Contact Centre

Requests for processing of sub-municipal data estimates from the 2011 Census	Year of request							
	2015	2016	2017	2018	2019	2020	2021	**Total**
No. of users who made at least one request with reference to the year of the one/last request (users eligible for the survey)	5	5	6	10	9	8	1	**44**
No. of "unreachable" eligible users	*1*	*1*	-	-	-	-	-	***2***
No. of users contacted for the survey	*4*	*4*	*6*	*10*	*9*	*8*	*1*	***42***
Number of users who participated in the survey (respondents)	3	2	6	7	5	8	1	**32**
Response rate (%)	75.0	50.0	100.0	70.0	55.6	100.0	100.0	**76.2**

Source: Istat - Directorate for Communication, Information and Services to Citizens and Users - Management and Dissemination Service

7 Analysis of responses

The answers provided by the users who took part in the survey were analysed from different perspectives to study them:

- the type of user, the reasons and purposes for requesting such specific processing;
- the degree of satisfaction with the level of spatial and classificatory detail of the data provided for the purposes for which they were requested;
- the enhancement of the data provided through GIS tools or integration with other sources or archives;
- the degree of satisfaction with the quality of the data received;
- the possibility of requesting new supplies of similar or different data in the future;
- the degree of satisfaction with the service received from the Istat Contact Centre.

7.1 Type of user, reason for request and purpose

Table 7.1 shows the data for the first two questions of the questionnaire: about the type of user, 62.5% belonged to the category "Research Institution, University, School", 28.1% to the sector "Enterprise, Self-employed" and only 9.4% to "Public Administration". Among the reasons for requesting data, 68.7% were for "Analysis and research", 21.9% for "Commercial purposes" and 9.4% for "Policy planning".

Table 7.1 - Distribution of respondents by "Type of user" and "Reason for request"

Type of user	Reason for request				Total	%
	Analysis and research	Commercial purposes	Policy planning	Other		
Research Institution, University, School	20	-	-	-	**20**	62.5
Enterprise, Self-employed	2	7	-	-	**9**	28.1
Public Administration	-	-	3	-	**3**	9.4
Other	-	-	-	-	**-**	-
Total	**22**	**7**	**3**	**-**	**32**	
%	68.7	21.9	9.4	-		**100.0**

Source: Authors' processing of survey results

The cross-analysis of the user type with the reason for the request shows a strong association between the two characteristics; finally, there is a strong concentration of cases (29 out of 32) amounting to 90.6% in the first two categories of user type and reason for the request.

Using the information collected with the third question, to which 30 users (93.7%) responded, it was possible to obtain an overview of the purposes for which users requested the provision of data. The prevailing purpose is to conduct spatial analysis, in different areas or for different purposes, to assess:

- socio-economic transformations;
- mobility in the territory;
- urban expansion and urban transformations;
- phenomena of social segregation (e.g. in schools);
- housing needs;
- energy needs for the definition of appropriate energy strategies.

The spatial analyses described by users also concerned more general public administration or commercial objectives (geo-marketing). Finally, in some cases, the data were used to carry out spatial classifications for specific objectives, including electoral ones.

7.2 Satisfaction of spatial and classificatory detail of data

This section analyses the answers to the fourth question on the degree of satisfaction with the level of informative, spatial and classificatory detail of the data provided, concerning the purposes for which they were requested (Table 7.2).

Table 7.2 - Distribution of respondents by "Reason for request" and "Level of satisfaction with the information detail of the data provided"

Reason for request	Satisfaction with the information detail of the data provided: Yes, fully	Yes, but they would have liked more detailed data	No	**Total**	%
Analysis and research	3	17	2	**22**	68.7
Commercial purposes	5	1	1	**7**	21.9
Policy planning	1	2	-	**3**	9.4
Total	**9**	**20**	**3**	**32**	
%	28.1	62.5	9.4		**100.0**

Source: Authors' processing of survey results

90.6% of the respondents were satisfied with the level of informative detail of the data received. In particular, 28.1% of users were fully satisfied, especially those who used the data for commercial purposes, while 62.5% would have liked more informative detail, especially those who requested the data for analysis or research purposes.

In particular, among respondents who were generally satisfied (29 out of 32) with the informative detail of the data received (including those who would have liked more detailed data), 69.0% used it for analysis or research purposes, 20.7% for commercial purposes, and 10.3% for policy planning.

Finally, 9.4% of users were dissatisfied with the information detail of the data received, in particular, due to the inadequate spatial level or the impossibility of cross-referencing the data received with other information.

7.3 Enhancement of provided data

The results of the analysis of the answers to the two questions concerning the use of data received with GIS[8] tools and integration with other statistical sources respectively are presented here. Table 7.3 shows the summary of the cross-referenced answers with the different types of users observed. Table 7.4 shows the results of the cross-referencing of the answers with the different reasons for which data were requested.

8 GIS (Geographic Information System) systems are computerised information systems that enable the acquisition, recording, analysis, visualisation, restitution, sharing and presentation of information derived from geographical data.

As a common datum between the two tables (shown in the last two rows), it can be noted that: 62.5% of the respondents used the data with GIS tools, while 37.5% did not use them in such systems; 65.6% of the users integrated the received data with data from statistical sources or administrative archives, while 34.4% did not use them in an integrated manner.

Table 7.3 - Distribution of respondents according to "Use of data with GIS tools" and "Integration with other sources" by "Type of user"

Type of user	Use of GIS tools		Integration of sources		Total	%
	Yes	No	Yes	No		
Research Institution, University, School	12	8	12	8	**20**	62.5
Enterprise, Self-employed	8	1	8	1	**9**	28.1
Public Administration	-	3	1	2	**3**	9.4
Other	20	12	21	11	**32**	
Total	**62.5**	**37.5**	**65.6**	**34.4**	**32**	
%	68.7	21.9	9.4	-		**100.0**

Source: Authors' processing of survey results

Table 7.4 - Distribution of respondents according to "Use of data with GIS tools" and "Integration with other sources" by "Reason for request"

Reason for request	Use of GIS tools		Integration of sources		Total	%
	Yes	No	Yes	No		
Analysis and research	14	8	13	9	**22**	68.7
Commercial purposes	6	1	7	-	**7**	21.9
Policy planning	-	3	1	2	**3**	9.4
Total	**20**	**12**	**21**	**11**	**32**	
%	62.5	37.5	65.6	34.4	**32**	**100.0**

Source: Authors' processing of survey results

Cross-referencing the answers with the "Type of user", it can be observed (Table 7.3) that among those who have used data with GIS tools (20 out of 32), 60% belong to the category "Research Institution, University, School" while the remaining 40% refer to the type "Enterprise, Self-employed"; no user from the group "Public Administration" declared having used data through GIS.

As regards the possibility of integrating the data with other administrative sources or archives, among those who stated that they had done so (21 out of 32), 57.1% belonged to the “Research Institution, University, School” sector, 38.1% to the “Enterprise, Self-Employed” context and the remaining 4.8% to the “Public Administration” context.

Cross-referencing the answers with the “Reason for request” (Table 7.4) shows that among those who used the data in GIS, 70% requested it for “Analysis and research” while the remaining 30% acquired it for “Commercial purposes”; none of those who used the data for “Policy planning” used it in GIS. As regards the integration of data with other sources or administrative archives, 61.9% requested them for “Analysis and research”, 33.3% for “Commercial purposes” and the remaining 4.8% for “Policy planning”.

Table 7.5 shows the results of cross-referencing the answers to the two questions concerning the use of data received with GIS tools and its integration with other sources. 53.1% of the interviewed users (17 out of 32) showed advanced specialisation in data processing, applying them either through GIS or in an integrated way with other data sources. 9.4% (3 out of 32) used them only through GIS and another 12.5% (4 out of 32) exclusively in an integrated manner with other data; the remaining 25% (8 out of 29) stated that they did not use the data either with GIS tools or in an integrated manner with other data.

Table 7.5 - Distribution of respondents according to “Use of data with GIS tools” and “Integration with other sources”

		Integration of sources		
		Yes	No	**Total**
Use of GIS tools	Yes	17	3	**20**
	No	4	8	**12**
	Total	**21**	**11**	**32**

Source: Authors’ processing of survey results

Returning now to the set of respondents who stated that they integrated the data received with other statistical sources (21 out of 32), in 19 cases the sources or archives used were specified. These are grouped as follows:

- Sources of economic data: Internal Revenue Service, Bank of Italy, Ministry of Economy and Finance, Chamber of Commerce;

- Institutional sources: Istat, Eurostat, Ministry of the Interior;
- Sources of data on housing and buildings: Topographic Databases, Land Registry, Military Geographic Institute (IGM), Real Estate Agencies;
- Sources of school data: Italian National Institute for the Evaluation of the Education and Training System (INVALSI);
- Internal or Local sources.

7.4 Data Quality Satisfaction

This section analyses the answers to the question on the degree of satisfaction with the level of quality[9] of the data provided about the purposes for which they were requested and acquired.

Table 7.6 shows that 53.1% of the respondents were “fully” satisfied with the quality of the data received, 40.6% were “only partly” satisfied, and 6.3% were negative about the quality of the data acquired.

Table 7.6 - Distribution of respondents on “Level of satisfaction with data quality” by “Reason for request”

Reason for request	Satisfaction level of data quality			Total	%
	Fully	Partially	No		
Analysis and research	9	12	1	**22**	68.7
Commercial purposes	5	1	1	**7**	21.9
Policy planning	3	0	0	**3**	9.4
Total	**17**	**13**	**2**	**32**	
%	53.1	40.6	6.3		**100.0**

Source: Authors’ processing of survey results

Cross-referencing the answers with the reason for the request, the users who used the data for analysis or research purposes are those who would have needed data characterised by a higher level of quality. This indicates a strong focus and expectation of academic and research users towards high-quality data, precisely because of the specificity of their study and analysis objectives.

9 In terms of relevance, accuracy, punctuality, clarity, comparability and consistency.

An analysis of the comments left by users who were partially satisfied with the quality of the data they received (12 out of 13) showed that rather than expressing an opinion on one of the aspects characterising the quality of the data, they specified difficulties or impossibilities in carrying out in-depth studies for all or some of the set study, analysis or research objectives. These hindrances were declared not because of the quality of the data, but for different reasons, including:

- unavailability of more detailed (informative or spatial) data;
- limited comparability of data with other sources or other census occasions for classification purposes.

7.5 Future requests

Let us now analyse what users have indicated about the possibility of making new requests for census data to Istat in the future.

Regarding the possibility of requesting new data from previous Population and Housing Censuses up to 2011, 78.1% responded that they would, while the remaining 21.9% did not exclude this possibility. None of the respondents stated that they would no longer request data from past censuses.

The next question referred instead to the possibility of requesting data from the new Permanent Census of Population and Housing in Italy. In particular, users were asked whether they would like to request the same census data supply as in the past or whether they would like to request data with a different classification or spatial detail. It should be noted that, as this was a multiple-choice question, several users gave more than one answer and the data had to be read in a different way than before.

From the results shown in Table 7.7, it can be observed that 50% of the respondents will confirm for the Permanent Census the same request made in the past or even other data, 43.8% will request, also or only, data referring to different details, 31.3% will request, also or only, data referring to different thematic crossings (other census topics). In addition, 12.5% of the respondents will request, also or only, other types of data, and finally 6.3% will not request any kind of data provision referring to the Permanent Census.

Table 7.7 - Number of observed answers to the question (multiple choice) whether the same or different data will be requested for the Permanent Census than in the past (separate answers)

Users' intentions to request the same or different data for the Permanent Census as for previous censuses		
Possible answers (multi-response question)	No. of responses	%
Yes, the same data supply	16	50.0
Yes, but data referring to different details	14	43.8
Yes, but data referring to different thematic crossings	10	31.3
Others	4	12.5
No, I will not ask for new data	2	6.3

Source: Authors' processing of survey results

As the user could provide more than one answer, these were reclassified (Table 7.8) to have a better indication of the possibility of future requests for Permanent Census data. In particular, 25% of the users will ask for the same data and nothing else, 40.6% will be interested in data with different information details, 25% will ask for both the same data as requested in the past and data for different details or census crossing, 3.1% will ask for other types of data without specifying the detail and, finally, 6.3% will ask for nothing. In summary, 93.7% will still ask for sub-municipal data concerning the Permanent Census of Population and Housing.

Table 7.8 - Number of observed answers to the question (multiple choice) whether the same or different data will be requested for the Permanent Census than in the past (re-classified responses)

Users' intentions to request the same or different data for the Permanent Census as for previous censuses		
Possible answers	No. of responses	%
It will only require the same data supply	8	25.0
It will only require data referring to different details or different thematic crossing	13	40.6
It will require the same supply of data and data referring to different details or different thematic crossings	8	25.0
It will require more data without specifying	1	3.1
It will not require new data	2	6.3
Total	**32**	**100.0**

Source: Authors' processing of survey results

Users belonging to the types "Research Institution, University, School" and "Enterprise, Self-employed" expressed a clear intention (around 90%) to request the same data supply and/or data with different details and/or different thematic crossings for the Permanent Census than in the past. This result indicates the need to ensure continuity in statistical production and the provision of census data with a high spatial detail. This is the only way to enable specialised users to continue their studies and research over the years and to guarantee a comparison with the past.

The analysis of the types of additional data that users are interested in obtaining from the Permanent Census shows that there is a strong need for data on enumeration areas and, in the case of dwellings and buildings, also georeferenced to the address. There is also a need for more integrated data, especially for dwellings and buildings.

Finally, several themes were indicated for which users expect to have data available with the Permanent Census. They include the occupation of the employed; commuting matrix between municipalities; type of system and type of fuel for heating and hot water; air-conditioning systems; and structural variables on unoccupied dwellings.

7.6 Contact Centre service satisfaction

In the last question of the questionnaire, the user was invited to express an opinion on the quality of the service received from the Istat Contact Centre. Tables 7.9 and 7.10 show the answers regarding the different degrees of satisfaction with the Istat service cross-referenced, respectively, with the "Type of user" and the "Reason for request".

As a datum in common between the two tables (reported in the last two rows) it should be noted that: 28.1% of respondents considered themselves fully satisfied with the service, 56.3% were satisfied, 15.6% were not very satisfied and no one declared themselves totally dissatisfied. It can therefore be assumed that 84.4% (27 out of 32) appreciated the service of the Istat Contact Centre at all stages of processing the customised processing, from the first contact to the sending of the data files. Around this general figure is the percentage of users who declared themselves satisfied depending on the type or reason for which the data were requested and used.

The cross-reference of the answers with the "Type of user" (Table 7.9) shows that 80% of the users belonging to the category "Research Institution, University, School" (16 out of 20), 88.9% of the type "Enterprise, Self-employed" (8 out of 9) and all users belonging to the category "Public Administration", were generally satisfied.

Table 7.9 - Distribution of respondents on "Level of service satisfaction" by "Type of user"

Type of user	Level of satisfaction with the Contact Centre service				Total	%
	Fully satisfied	Satisfied	Not very satisfied	Totally dissatisfied		
Research Institution, University, School	5	11	4	-	**20**	62.5
Enterprise, Self-employed	4	4	1	-	**9**	28.1
Public Administration	-	3	-	-	**3**	9.4
Total	**9**	**18**	**5**	**-**	**32**	
%	28.1	56.3	15.6	-		**100.0**

Source: Authors' processing of survey results

On the other hand, cross-referring the answers with the "Reason for request" (Table 7.10), 81.8% of the users who requested the data for "Analysis and research" (18 out of 22), 85.7% of those who requested it for "Commercial purposes" (6 out of 7) and all of those who used it for "Policy planning" were satisfied.

Table 7.10 - Distribution of respondents on "Level of satisfaction with the service" by "Reason for request"

Reason for request	Level of satisfaction with the Contact Centre service				Total	%
	Fully satisfied	Satisfied	Not very satisfied	Totally dissatisfied		
Analysis and research	6	12	4	-	**22**	68.7
Commercial purposes	3	3	1	-	**7**	21.9
Policy planning	-	3	-	-	**3**	9.4
Total	**9**	**18**	**5**	**-**	**32**	
%	28.1	56.3	15.6	-		**100.0**

Source: Authors' processing of survey results

At the end of the questionnaire, the respondents were allowed to point out any critical issues they had encountered or make suggestions for improving the data delivery and service of the Istat Contact Centre. Below is a summary of the comments left by the respondents (16 out of 32) grouped according to three different areas.

- Indications of supplies for the future:
  - structural data on housing and buildings;
  - data on non-residential buildings;
  - expansion of the information collected by the Permanent Census.

- Criticalities of the Contact Centre service:
  - excessively long time for the release of the supply;
  - excessive bureaucracy in practice;
  - high cost.
- Proposals for improving the service:
  - greater accessibility of micro-data;
  - possibility to interact with the technical services of Istat that produced the data;
  - lowering costs for students, Ph.D. students, and young people.

8. Analysis of the continuity of demand for census information

In this area, a study was conducted on the expected level of "continuity" of future requests, in terms of thematic content, compared to the past, through analysis of census crossings requested by users in the past and based on responses to the questionnaire on the possibility of requesting Permanent Census data.

The first step involved a careful review of all requests for customised processing of 2011 Census data, received by the Contact Centre from 2015 until early 2021, for the 44 users who received the data (see paragraph 5.2). The census crossings requested were subsequently classified by different census topics. Following this, for each of the 44 users considered in the analysis, the different types of data requested were identified (in many cases users requested data for different topics), also grouping those requests associated with the same user that arrived in different years during the period considered (2015-2021).

In this way, it was possible to assess, for each census topic, what proportion of the total number of users who acquired the data was requested.

The same analysis was also conducted on the 32 users who answered the questionnaire to assess which types of data those who took part in the survey requested. For these users, question 9 of the questionnaire defined the topic of data they could request for the Permanent Census, possibly with a different detail. This information, cross-referenced with that relating to the census topics of data requested in the past, gave useful indications as to which topics will have a greater propensity to be requested from Istat in the coming years, as the Permanent Census consolidates its data production and dissemination process. The results of the analysis are summarised in Table 8.1.

Taking census topics F - "Structural characteristics of dwellings" as an example, the following information can be read in Table 8.1:

- 31.8% of users who made requests to the Istat Contact Centre (14 out of 44) from 2015 to early 2021 asked for supplies that included data on the "structural characteristics of dwellings";
- 31.3% of the users who participated in the survey (10 out of 32) had requested data on the "structural characteristics of dwellings" in the past;

- 71.4% of the users who acquired data on the "structural characteristics of the dwellings" took part in the survey (10 out of 14);
- 90% of respondents who have asked for data on the "structural characteristics of dwellings" in the past (9 out of 10) will also ask for it concerning the Permanent Census.

Table 8.1 - Number of total Contact Centre users and survey respondents classified by the different types of census data asked in the past and to be asked again in the future

	Census topics	Contact Centre users	% of total users (44)	Interviewed users	% of total respondents (32)	Response rate of respondents by topic	Respondents requesting data on the same topic	Confirmation rate of the census topic
A	Education	7	15.9	6	18.8	85,7	2	33.3
B	Current activity status	8	18.2	5	15.6	62,5	3	60.0
C	Employment characteristics	25	56.8	20	62.5	80,0	15	75.0
D	Family type	7	15.9	5	15.6	71,4	3	60.0
E	Ownership/right of use of housing	6	13.6	4	12.5	66,7	3	75.0
F	Structural characteristics of dwellings	14	31.8	10	31.3	71,4	9	90.0
G	Heating systems and fuel	14	31.8	10	31.3	71,4	7	70.0
H	F and G	8	18.2	6	18.8	75,0	5	83.3
I	F or G	20	45.5	14	43.8	70,0	11	78.6
L	Buildings	3	6.8	2	6.3	66,7	1	50.0
M	Foreigners	8	18.2	8	25.0	100,0	7	87.5
N	Commuting	1	2.3	1	3.1	100,0	1	100.0

Source: Authors' processing of survey results

9. Concluding remarks

The survey described in this article was the first conducted by Istat on a specific target of users, intending to assess the match between user demand and the supply of reliable and timely statistics, according to principles of the European Statistics Code of Practice. An attempt was made to measure the extent to which essential information for different actors in the economy, research and institutions is met by that produced by Istat, whose main vocation is to produce and disseminate data and analysis on relevant phenomena in an accessible and clear manner.

The evaluation was equally extended to the user support service offered by the Istat Contact Centre which, in the extensive and articulated data dissemination system, represents the main channel for requesting personalised processing.

The results of the survey provide a well-established positive opinion among specialist users, many of whom have become loyal over time, on which to draw for further reflections/planning between the production and dissemination of statistical information.

Moreover, the results of the survey made it possible to define the set of census data referring to the sub-municipal levels that are highly expected by users to conduct studies and research, also with the use of GIS and integration with other data sources. The data most frequently requested - and which will continue to be requested by users - concern the population’s level of education, current activity status, status in employment (position in the occupation), type of work (occupation) or sector of economic activity (industry) of employed, type of family and structural characteristics of dwellings.

It follows that Istat must concentrate more effort, both in the data collection phase and in the phase of defining statistical methodologies, so that the supply of sub-municipal census data for the new Permanent Census is relevant to users’ needs.

As we have not found in other foreign countries a user survey similar to the one presented in this article, we suggest that this type of comparison with specialised users (the census stakeholders) and analysis of user needs should be done in advance by all NSI of countries that intend to innovate their census strategy in favour of greater integration of administrative and sample data, to properly assess the impact on data supply concerning relevance to user expectations.

Annex – Istat's user Satisfaction Questionnaire

This section describes the questions and possible answers of the questionnaire used by Istat for the survey presented in this article.

1. At the time of your request, what type of user did you belong to?

- ○ Research Institution, University, School
- ○ Enterprise, Self-employed
- ○ Public Administration
- ○ Other (specify)

2. Can you indicate the main reason for your request for data?

- ○ Analysis and research
- ○ Commercial purposes
- ○ Policy Planning
- ○ Other (specify)

3. Can you briefly explain the objectives for which you used the data provision?

   Short description of the topic of the analysis/research or presentation of the abstract of the work carried out

4. Did the level of spatial detail as well as the classificatory level (*e.g.* 5-year or 10-year "age class"; 6-mode or 17-mode "level of education"; 3-mode or 21-mode "sector of economic activity") of the data provided meet your needs?

- o Yes, fully
- o Yes, but I would have had more opportunities with more detailed data
- o No (specify)

5. Has the data received been used with GIS tools?

- o Yes
- o No

6. Has the data received been integrated with other statistical sources?

- o Yes (please specify)
- o No

7. Did the level of quality (in terms of relevance, accuracy, timeliness, clarity, comparability and consistency) of the data provided meet your needs?

- o Yes, fully
- o Only partly, because I would have needed data with a higher level of quality (please specify the reasons)
- o No (specify reasons)

8. Do you think that in the future you will make new requests for the supply of census data to Istat, also about Censuses other than that of 2011?

- o Yes
- o No
- o I don't know, but I don't rule it out

9. Concerning the new Permanent Population and Housing Census, do you think that in the future you will be able to request the same type of data as in the past or will you need to request data with a different classification or spatial detail? [Multiple answers are possible].

- o Yes, the same data supply
- o Yes, but data referring to different details
- o Yes, but data referring to different census topics
- o Yes, but other data (please specify which types of data you may require)
- o No, I will not ask for new data

10. How would you rate the overall quality of the service you receive from Istat?

- o Fully satisfied
- o Satisfied
- o Not very satisfied
- o Totally dissatisfied (specify why)

11. If you wish, you can leave a comment, point out any critical issues or make suggestions to improve the service.

References

Borrelli, F., G. Carbonetti, S. Dardanelli, L. De Felici, E. Fiorello, M. Marrone, e M. Verrascina. 2012. "La progettazione dei censimenti generali 2010-2011: misure di accuratezza di tavole di diffusione per livelli territoriali e dettagli informativi". *Istat working papers*, N. 15/2012. Roma, Italia: Istat. https://www.istat.it/it/files//2012/12/Istat-Working-Papers-n.-15-2012.pdf.

Borrelli, F., G. Carbonetti, L. De Felici, E. Fiorello, e M. Marrone. 2011. "La progettazione dei censimenti generali 2010-2011: disegni campionari e stima di errori di campionamento". *Istat working papers*, N. 2/2011. Roma, Italia: Istat. https://www.istat.it/it/files/2011/06/Istat_Working_Papers_2_2011.pdf.

Carbonetti, G. 2009. "Use of Sampling Strategy in the Italian Population Census and Accuracy of Estimates for Different Territorial Domains". Paper presented at the *First Italian Conference on Survey Methodology (ITACOSM09)*. Siena, Italy, 10-12 June 2009.

Carbonetti, G., A. Ciccarese, and R. Roncati. 2022. "User satisfaction survey of sub-municipal data from the 2011 Italian Population and Housing Census". Paper presented at the *European Conference on Quality in Official Statistics (Q2022)*. Vilnius, Lithuania, 8-10 June 2022.

Carbonetti, G., and M. Fortini. 2008. "Sample results expected accuracy in the Italian population and housing census". Note by the Italian National Institute of Statistics presented at the *Joint UNECE/Eurostat Meeting on Population and Housing Censuses*. Eleventh Meeting, Geneva, Switzerland, 13-15 May 2008. https://unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.41/2008/4.e.pdf.

Carbonetti, G., M. Fortini, and F. Solari. 2008. "Innovations on methods and survey process for the 2011 Italian population census". Paper presented at the *European Conference on Quality in Official Statistics (Q2008)*. Roma, Italy, 9-11 July 2008.

Carbonetti, G., and M. Verrascina. 2010. "Accuracy Evaluation of Dissemination Data Adopting a Sampling Strategy in the 2011 Italian Population Census". Paper presented at the *European Conference on Quality in Official Statistics (Q2010)*. Helsinki, Finland, 4-6 May 2010.

Central Statistics Office - CSO. 2022. *Press Statement CSO Customer Satisfaction Survey 2022*. Cork, Ireland: CSO (last accessed in March 2023). https://www.cso.ie/en/csolatestnews/pressreleases/2022pressreleases/pressstatementcustomersurvey2022/.

European Statistical System - ESS, and Eurostat. 2017. *European Statistics Code of Practice. For the National Statistical Authorities and Eurostat* (*EU statistical authority*). Luxembourg: Publication Office of the European Union. https://ec.europa.eu/eurostat/web/quality/europeanquality-standards/european-statistics-code-of-practice.

Falorsi, S. 2017. "Census and Social Surveys Integrated System". Note by the Italian National Institute of Statistics presented at the *Joint UNECE/Eurostat Meeting on Population and Housing Censuses*. Nineteenth Meeting, Geneva, Switzerland, 4-6 October 2017. https://unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.41/2017/Meeting-Geneva-Oct/WP23_ENG.pdf.

Ghana Statistical Service - GSS. 2023. *User Satisfaction Survey*. Accra, Ghana: GSS (last accessed in March 2023). https://statsghana.gov.gh/gssmain/fileUpload/pressrelease/2022%20User%20Satisfaction%20Survey_Abridged%20Report_21.02.23%20-%20Final_online.pdf.

Hellenic Statistical Authority - ELSTAT. 2023. *User Satisfaction Survey*. Piraeus, Greece: ELSTAT (last accessed in March 2023). https://www.statistics.gr/el/user-satisfaction-survey.

Institute of Statistics - INSTAT. 2023. *User Satisfaction Survey*. Tirana, Albania: INSTAT (last accessed in March 2023). https://www.instat.gov.al/en/documentation/quality-in-statistics/user-satisfaction-survey/.

Istituto Nazionale di Statistica - Istat, Directorate for communication, information and services to citizens and users - DCCI. 2022. "Service charter for dissemination and communication". In Istat, DCCI. *Istat Statistical Dissemination System*. Roma, Italy: Istat. https://www.istat.it/it/files//2011/03/Istat-statistical-dissemination-system.pdf.

Istituto Nazionale di Statistica - Istat, Directorate for communication, information and services to citizens and users - DCCI. 2021. *The Contact Centre Guidelines for Assistance in Data Retrieval and Customised Processing*. Roma, Italy: Istat.

Legge 27 dicembre 2017, n. 205. “Bilancio di previsione dello Stato per l’anno finanziario 2018 e bilancio pluriennale per il triennio 2018-2020”. *Gazzetta Ufficiale Serie Generale*, N. 302 del 29 dicembre 2017, Supplemento ordinario N. 62. https://www.gazzettaufficiale.it/eli/id/2017/12/29/17G00222/sg.

National Bureau of Statistics of Tanzania - NBS, and Office of the Chief Government Statistician in Zanzibar - OCGE. 2014. *The Statistics User Satisfaction Survey*. Dodoma, Tanzania: NBS (last accessed in March 2023). https://www.nbs.go.tz/nbs/Dataset/USS2014/NBS%20Satisfaction%20Survey%20Report%202014%20.pdf.

National Institute of Statistics of Rwanda - NSIR. 2019. *User Satisfaction*. Kigali, Rwanda: NSIR (last accessed in March 2023). https://www.statistics.gov.rw/publication/user-satisfaction-survey20182019.

National Statistical Institute of Bulgaria - NSI. 2023. *Questionnaire for the survey of user satisfaction with the offered statistical data products and services*. Sofia, Bulgaria: NSI (last accessed in March 2023). https://www.nsi.bg/en/content/16537/questionnaire-survey-user-satisfaction-offered-statistical-data-products-and-services.

Office for National Statistics - ONS. 2016. *ONS Customer Satisfaction Survey*. London, UK: ONS (last accessed in March 2023). https://www.ons.gov.uk/aboutus/whatwedo/statistics/consultationsandsurveys/allconsultationsandsurveys/onscustomersatisfactionsurvey.

Presidenza del Consiglio dei Ministri, Dipartimento della Funzione Pubblica. 2004. *Direttiva del Ministro della Funzione Pubblica sulla Rilevazione della qualità percepita dai cittadini*. Roma, Italia: Presidenza del Consiglio dei Ministri.

Spanish Statistical Office - INE 2023. *User survey*. Madrid, Spain: INE (last accessed in March 2023). https://www.ine.es/ss/Satellite?c=Page&cid=1259944133654&pagename=MetodologiaYEstandares%2FINELayout&L=0.

Statistical Office of the Republic of Slovenia - SURS. 2022. *User satisfaction survey*. Ljubljana, Slovenia: SURS (last accessed in March 2023). https://www.stat.si/StatWeb/en/News/Index/10682.

Statistics Canada. 2023. *Statistics Canada Client Survey 2021*. Ottawa, Ontario, Canada: Statistics Canada (last accessed in March 2023). https://www.statcan.gc.ca/en/consultation/2021/sccs2020.

Statistics Estonia. 2023. *User Surveys*. Tallinn, Estonia: Statistics Estonia (last accessed in March 2023). https://www.stat.ee/en/statistics-estonia/about-us/user-surveys.

Statistical Office of the Republic of Serbia - PBC. 2021. *Results of the User Satisfaction Survey*. Belgrade, Serbia: PBC (last accessed in March 2023). https://www.stat.gov.rs/en-us/vesti/20211215-istrazivanje-o-zadovoljstvu-korisnika/?a=0&s=2501.

Statistical Service of Cyprus - CYSTAT. 2021. *User Satisfaction Survey*. Nicosia, Cyprus: CYSTAT (last accessed in March 2023). https://www.cystat.gov.cy/en/StaticPage?id=1144.

Credit constraints and firm characteristics. An empirical assessment of a survey-based Istat's indicator

Antonio Ughi[1], Alessandro Zeli[2], Silvia Lombardi[2]

Abstract

This paper investigates the relationship between credit constraints and firm characteristics in Italy during 2015-2020. We derive an annual measure of perceived credit constraints using qualitative information from the Istat Business Confidence Survey and relate it to firm financials and other characteristics from the Italian Business Register. Using a linear probability model, we test the informative power of the indicator and find higher constraints for smaller and less productive firms located in Southern Italian regions. In addition, we explore the relationship between the new indicator of perceived credit constraint and financial conditions, finding that financially healthier firms experience lower obstacles in accessing external credit. The analysis brings novel empirical evidence about perceived measures of credit constraints in Italy over several years and provides a helpful indicator for future firm-level empirical and policy studies.

Keywords: Credit constraints, firm dynamics, Italian economy.

JEL Classi ication: O16, G20.

DOI: https://doi.org/10.1481/ISTATRIVISTASTATISTICAUFFICIALE_1-2-3.2023.03

---

1 Antonio Ughi (antonio.ughi@santannapisa.it), Sant'Anna School of Advanced Studies, Pisa, Italy.
2 Alessandro Zeli (zeli@istat.it); Silvia Lombardi (lombardi@istat.it), Italian National Institute of Statistics – Istat.

The views and opinions expressed are those of the authors and do not necessarily reflect the official policy or position of the Italian National Institute of Statistics - Istat.

The authors would like to thank the anonymous reviewers for their comments and suggestions, which enhanced the quality of this article.

1. Introduction

Accessing external finance is a crucial factor for firms. The literature has shown that limited access to external finance negatively influences firms' R&D and innovation activities (Hall 2002; Brown *et al.* 2009), wages (Michelacci *et al.* 2009), export (Minetti *et al.* 2011; Secchi *et al.* 2016), survival probability (Musso *et al.* 2008), default risk (Bottazzi *et al.* 2011), investments (Fazzari *et al.* 1988; Kaplan *et al.* 1997, 2000; Almeida *et al.* 2007), and growth (Bottazzi *et al.* 2014)[3].

Firms may become financially constrained if their activities are limited by the difficulty of obtaining external funds, e.g. because of information asymmetries between lenders and borrowers (Stiglitz *et al.* 1981; Jaffee *et al.* 1990). In turn, asymmetries may be related both to factors internal and external to firms. On the one hand, e.g. lack of proper information may increase perceived riskiness and induce a lack of financing for Small and Medium Enterprises (SMEs) (Cosh *et al.* 2009; Storey 2016). Indeed, SMEs are usually unlisted, may have less transparent track records, no collateral, and carry out activities which are more difficult to evaluate *vis-à-vis* larger firms – resulting in higher costs of external funds (Berger *et al.* 2006; Revest *et al.* 2012). Relatedly, newly established and young firms usually have a limited credit history, which makes it difficult for banks to predict the future probability of loan repayment and reduce financing opportunities (Gertler 1988; Devereux *et al.* 1990; Beck *et al.* 2006). Similarly, other contributions suggest that access to finance depends on the strength of firms' balance sheets, e.g. better financial figures reduce perceived riskiness (Jimenez *et al.* 2012; Holton *et al.* 2013; McQuinn 2019). On the other hand, firms may experience worse financing conditions for other external factors, for example, being in regions where financial markets are less developed (Djankov *et al.* 2007). In addition, lower financing opportunities may be present because of specific sectoral dynamics, as well as for the overall stance of the business cycle (Blanco *et al.* 2018; Finnegan *et al.* 2023).

3 The relation among firm performance, one the one hand, and financial constraints, on the other, may not be unidirectional. For example, Lahr and Mina (2020) show that innovation – and in particular new-to-market and product innovations – increases the likelihood for firms of facing financial constraints, due to uncertain market outcomes associated to innovative activities.

Measuring financial constraints poses challenges, as this information is typically unobserved, unlike other financial variables. Several contributions have tried to estimate measures of financial constraints indirectly. These methodologies include e.g. investment–cash flow sensitivities (FHP) (Fazzari *et al.* 1988) and the Kaplan and Zingales approach (KZ) (Kaplan *et al.* 1997)[4]. The seminal approach of Fazzari *et al.* (1988)[5] mainly suggests that financing constraints can be identified using the sensitivity of firm investment to internal funds (cash flow). Indeed, financially constrained firms, due to limited access to external financing, rely more heavily on internally generated funds to finance their investments. As a result, investment spending of constrained firms becomes more sensitive to changes in their cash flow positions. Kaplan and Zingales (1997, 2000) question the validity of this interpretation and show that, under certain assumptions, investment–cash flow sensitivities may increase as financing constraints are relaxed – with similar results found later by Almeida and Campello (2007). Among other concerns[6], they argue that financial variables employed in the FHP approach may become endogenous and not have a straightforward relation to constraints (Farre-Mensa *et al.* 2016).

In general, considerable debate exists concerning the possible methods for measuring financial constraints, as each method relies on diverse theoretical and/or empirical assumptions (Hadlock *et al.* 2010). In turn, another strand of literature has used measures of financial constraints collected through firm surveys. Indeed, surveys would allow assessing the role and factors influencing financing obstacles more directly, eliminating the need to infer constraints from financial information (Beck *et al.* 2006). Although this approach has its limitations – such as potential biases in self-reported data and representativeness concerns – the empirical literature using survey-based indicators generally agrees on the factors linked to credit constraints. For example, Beck *et al.* (2006) use the "World Business Environment Survey" (WBES) data[7] on a sample of over 10,000 firms from 80 countries and investigate the determinants of financing obstacles for firms. The authors find

4 Other approaches include the Whited and Wu index of constraints (WW) (Whited *et al.* 2006) and the Hadlock-Pierce index of constraints (Hadlock *et al.* 2010). The former derives a measure of constraints from a structural model, while the latter uses financial fillings to categorise financial constraints.

5 For a review of related works building on this approach, see Farre-Mensa and Ljungqvist (2016).

6 For an extensive literature review on the estimation of financial constraints, see also Hadlock and Pierce (2010).

7 The WBES is a major firm-level survey managed by the World Bank. The authors use data for 1999 and 2000 in developing and developed countries.

that older, larger, and foreign-owned firms report lower levels of financial constraints and that higher institutional development is associated with better access to finance. Similarly, Ferrando and Griesshaber (2011) document higher financial constraints for, e.g. younger EU firms, using self-assessment measurements from the "Survey on the Access to Finance of small and medium-sized Enterprises" (SAFE)[8,9]. Canton *et al.* (2013) investigate the determinants of perceived bank loan accessibility at the firm level using survey data for nearly 3,500 SMEs in 25 EU countries. They find that, overall, the youngest and smallest SMEs have the worst perception of access to loans[10]. Finally, Kuntchev *et al.* (2016) also confirm the above-mentioned relations using another survey-based measure of financial constraints (from the World Bank "Enterprise Survey").

Our analysis builds on and contributes to this literature, especially the one that derives and tests direct measures of financial constraints from survey data. Notably, this work assesses the relation among self-reported measures of credit constraints and firm characteristics for a large sample of Italian firms over a long time. We derive an indicator of perceived credit constraint at the firm level relying on qualitative information from the Istat "Business Confidence Survey", uniform across sectors and years. Using a linear probability model, we ask which firm-level characteristics relate to more significant perceived obstacles in accessing external finance – testing our indicator with results from the literature. We find higher constraints for smaller and less productive firms in less developed Italian regions. In addition, we explore the relationship between our new indicator and financial conditions at the firm level, finding that better financial conditions are related to lower perceived obstacles in accessing external finance. In particular, higher profitability appears associated with lower constraints, while higher debt levels are associated with higher constraints.

8 The authors use the second wave of the ECB-European Commission "Survey on the Access to Finance of small and medium-sized Enterprises" (SAFE), which provides evidence on the financial situation, financing needs and the access to financing of small and medium-sized enterprises as well as of a comparable sample of large firms in the euro area during the second half of 2009.

9 The authors find mixed results for the size variable while they find a significant relationship between financial constraints and ownership structure (e.g. listed firms are less likely to be financially constrained).

10 Looking at the effects of being financially constrained, Ferrando and Mulier (2022) use survey data from the "Survey on the access to finance of enterprises" (SAFE) for 9 EU countries between 2010 and 2014, and analyse the effect of being a discouraged borrower (i.e. firms that do not apply for a bank loan because they fear that their application will be rejected). They report a strong negative correlation between discouragement and firm investment and growth. Gómez (2019) performs a related analysis, using the SAFE survey for 12 EU countries from 2014 to 2017, and finds that credit constraints have strong negative effects on investment in fixed assets.

The remainder of the work is organised as follows. Section 2 describes the data sources, variables – in particular, the derivation of our annual measure of perceived credit constraints – and shows descriptive statistics. Section 0 focusses on the empirical strategy and Section 5 on the main estimation results and robustness checks. Section 6 concludes and elaborates on the avenues for further research.

2. Data and perceived credit constraint indicator

2.1 Data sources

To perform the analyses, we integrate three different sets of data over the period 2015-2020. To derive information on perceived credit constraints, we rely on qualitative data from the Istat “Business Confidence Survey” (BCS). This survey collects uniform quarterly information on the economic sentiment of enterprises active in manufacturing, construction, retail, and other services sectors. Secondly, we use administrative financial data on income statements and balance sheet accounts of Italian corporations drawn from official records filed in the Italian Chambers of Commerce (CCs) and available in Istat. Finally, we use structural variables (i.e. employment, economic activity, localisation) from the Italian Business Register (BR), which provides a framework for the integration of the above-mentioned microdata.

As far as the sample survey information is concerned, firms are asked – each quarter – to provide information on whether they perceive access to credit as ‘more favourable’, ‘constant’, ‘less favourable’, or ‘unknown’ – concerning the previous quarter. The survey collects harmonised cross-sectoral data and overall, in the period 2015-2020, it allows to retrieve more than 170,000 records (quarterly microdata), corresponding to ca. 14,000 enterprises.

Administrative data on annual economic accounts and income statements allow us to compute the main financial indicators used for the analysis. This data source involves about 700,000 corporations each year, corresponding to almost 4 million observations across 2015-2020. This data source mainly refers to limited companies that account for 22,5% of total firms and 57,2% of employment (in 2020).

Finally, firm annual financial data are complemented with structural information from the Business Register, including firm economic (NACE Rev.2) sector of activity, geographical location, and employment.

For firms with balance sheet information, data integration results in 87,883 complete quarterly records for the period 2015-2020 (Table 2.1). Apart from 2015, for which only the fourth quarter is available, each year has more than 15,000 quarterly records – related to about 5,000 firms yearly.

Table 2.1 - Quarterly records and number of firms (a)

Year	N. of records	N. of firms
2015	4,059	4,059
2016	19,820	7,131
2017	14,857	4,406
2018	16,855	5,226
2019	17,120	5,164
2020	15,172	4,593
Total	**87,883**	**30,579**

Source: Authors' elaborations on BCS-BR-CCs databases
(a) Only the fourth quarter of 2015 is available for the analysis.

Overall, integrated data refer to almost 9,000 distinct firms (Table 2.2, first column) whose average quarterly presence is 9.9 times. Each firm is present at least one quarter in each year, while no firms are present each quarter for all years (the maximum presence is 21 times). In comparison with the survey sample (Table 2.2, second column), integrated data have an overall satisfying share of coverage (Table 2.2, third column).

Table 2.2 - Number of univocal firms by macro sector

Macro sector (NACE Rev.2)	N. of univocal firms of integrated sample	N. of univocal firms of the survey sample	% Share of coverage of integrated sample
1 - Industry	5,132	6,784	75.6
2 - Construction	916	1,694	54.1
3 - Wholesale and retail trade	747	2,023	36.9
4 - Services	2,124	3,878	54.8
Total	**8,919**	**14,379**	**62.0**

Source: Authors' elaborations on BCS-BR-CCs databases

The missing companies (i.e. the companies not included in the panel) are mostly non-limited liability companies that are not obliged to submit their financial reports to the CCs.

As financial indicators and other firm characteristics are available annually, in the next subsection we compute a yearly firm-level aggregation of the quarterly variable about perceived credit constraints.

2.2 Perceived credit constraint indicator

Our main variable of interest is the firm-level measure of perceived credit constraints – derived from the 'BCS' survey. As discussed above, firms may report, each quarter, their financing (opportunities) conditions as 'more favourable', 'constant', 'less favourable', or 'unknown' – concerning the previous quarter (Appendix B reports the wording of the survey questions used to identify credit constraints)[11]. To relate this information to yearly firm characteristics, we retrieve an annual aggregation of such measures.

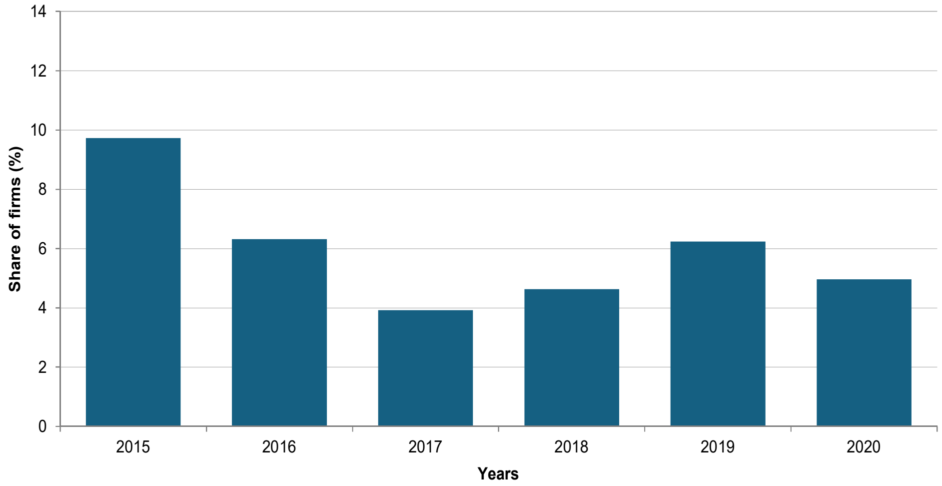
In each quarter, the survey information captures firms' perceived 'access to credit' *status* Q-o-Q. This implies that – absent reporting on firm-level initial conditions – the survey allows to derive only a relative status each quarter and not the actual levels. Moreover, since not all the firms report their conditions each quarter (see previous subsection), this creates missing data points in the firm-level status time series. Given these limitations, we derive an annual measure of perceived credit constraints proceeding in two steps. We first compute the number of times each firm reports 'less favourable' conditions in accessing credit relative to the number of quarters available each year (excluding quarters in which firms report an 'unknown' condition). Second, we define a firm as "credit constrained" – each year – if the condition 'less favourable' is reported more than half of the times over the quarters available for each firm (we test the robustness of such indicator also using other thresholds and restrictions on available quarters – see Section 5.1 and the Appendix A). We obtain a dummy indicator capturing perceived credit constraints at the firm level, which we employ for our empirical analysis. We start with a descriptive assessment of the indicator, which we present in the next Section.

11 Given our research question, we focus on Q.43. Other questions proved less informative – also given the lower firm response rates.

3. Descriptive evidence

Figure 3.1 displays the annual share of firms reporting perceived credit constraints using the indicator derived in the previous Section. As shown, the share of constrained firms appears at its maximum in 2015, decreases over 2016 and 2017, and increases again in 2018 and 2019 – finally decreasing in 2020. This dynamic seems consistent with the business cycle conditions observed in Italy over the period (Istat 2023; Istat 2021). Indeed, starting from 2015, and over 2016 and 2017, Italy experienced a recovery in GDP growth and a general ease in credit conditions. On the contrary, 2018 and 2019 saw a weakening of GDP growth, higher average interest rates, and declining credit and loan supply to business firms, with likely impacts on firm performance (Istat 2021). Regarding 2020, the lower share of constrained firms – despite the COVID-19 outbreak – may be a result of the increase in bank loans (Istat 2023) and the extensive number of support measures implemented following the pandemic (for instance, debt guarantees or loans *moratoria*).

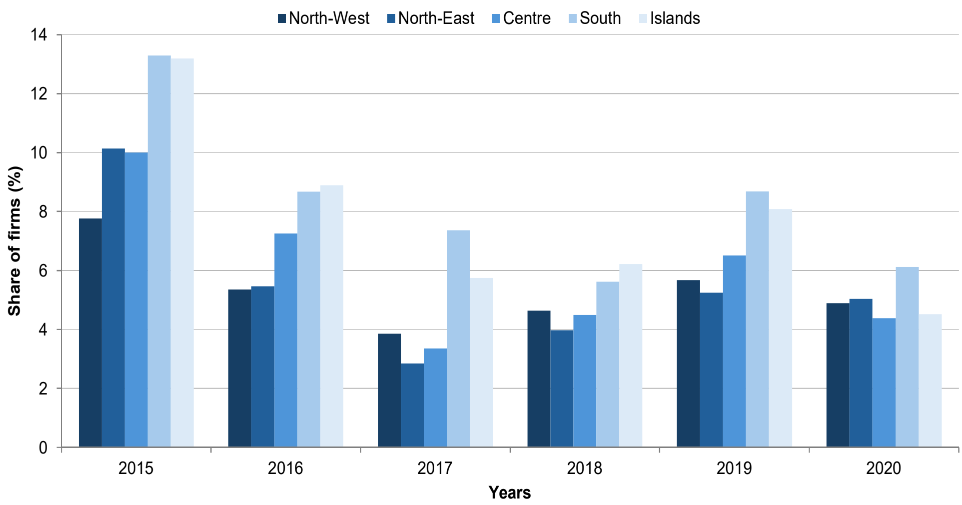
Figure 3.1 - Yearly share of firms reporting perceived credit constraints. Years 2015-2020



Source: Authors' elaborations on BCS-BR-CCs databases

Further investigating the yearly shares of firms reporting credit constraints, Figure 3.2 shows statistics by macroregion. While the yearly trends follow those reported in Figure 3.1, there seems to be considerable heterogeneity across Italian regions. Indeed, the Southern part of the country (South and Islands) has the highest share of constrained firms, while the North-West, North-East, and Centre regions display lower shares. This evidence seems in line with other studies pointing to more adverse financing conditions in less (financially) developed regions, as is the case for the South of Italy (Istat 2021).

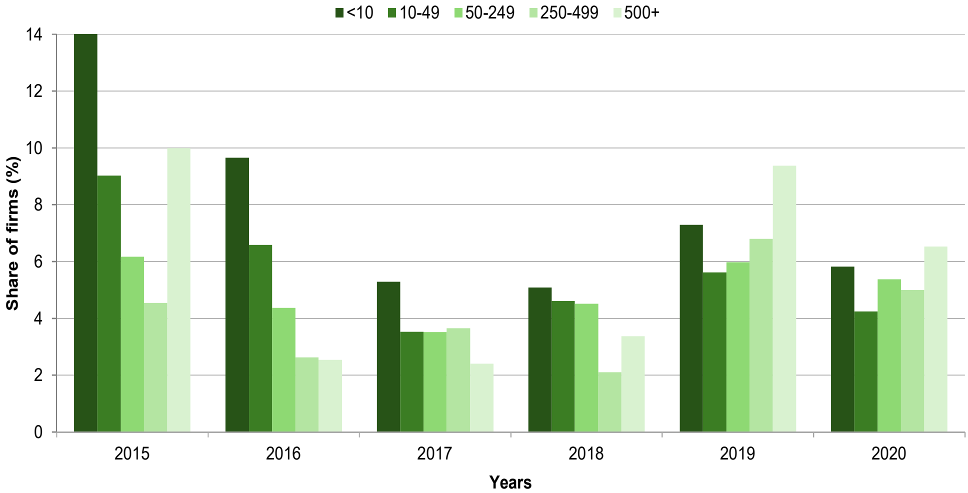
Figure 3.2 - Yearly share of firms reporting perceived credit constraints by macroregion. Years 2015-2020 (a)



Source: Authors' elaborations on BCS-BR-CCs databases
(a) Macroregion refers to the NUTS-1 regional classification.

As discussed in Section 1, Small and Medium Enterprises (SMEs) are usually the most affected by credit constraints. At least to some extent, this seems to be the case in Italy over the period considered, as shown in Figure 3.3. Indeed, there appears to be a negative relationship between perceived credit constraints and size for 2015-2018 (the years associated with a relatively calm business cycle period). For 2016 and 2017, the relation appears monotonically decreasing with size, while overall decreasing shares characterise 2015 and 2018 except for the 500+ class.

Figure 3.3 - Yearly share of firms reporting perceived credit constraints by firm size class. Years 2015-2020



Source: Authors' elaborations on BCS-BR-CCs databases

If one looks at the relation between perceived credit constraints and productivity[12] (Figure 3.4), the picture appears relatively more in line with other findings from the literature. Indeed, the share of firms reporting credit constraints appears higher for low productivity quantiles (especially the bottom 10% and 10%-30% of the productivity distribution). On the contrary, higher productivity quantiles seem on average associated with lower perceived constraints[13].

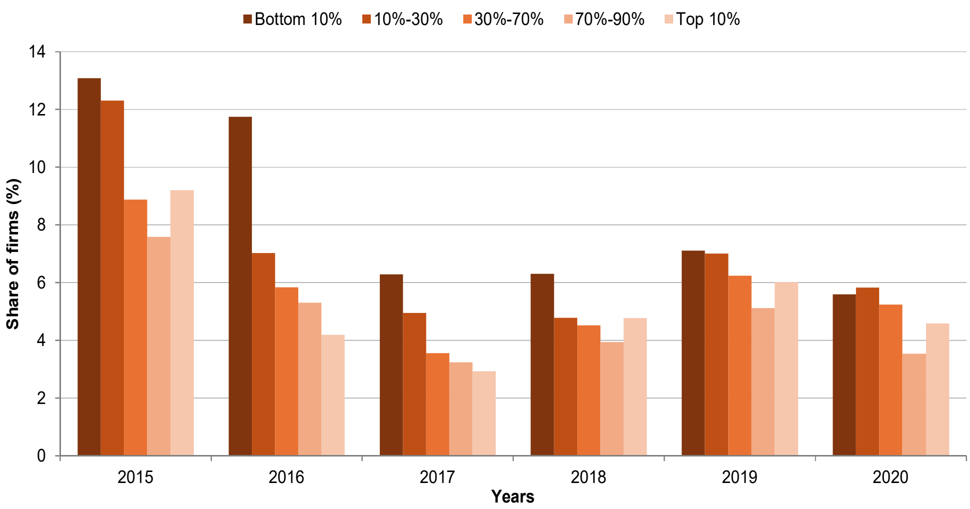
For the following analysis, we employ the above-mentioned variables (macroregion, size, and productivity) as well as a set of continuous financial indicators. In particular, to study the relations between perceived credit constraints and firm financial characteristics, we rely on a set of profitability (ROA, ROE), solvency (debt over assets), and liquidity (Current ratio) indicators. These may well be related to obstacles in access to credit. Indeed,

12 The nominal labour productivity variable is computed as value added over employees. Productivity classes are computed within sector-region (Figure 3.4).

13 For the top 10% productivity quantile, results show higher shares of firms' vis-à-vis, e.g. the 70%-90% quantile of the productivity distribution. This finding seems to align to the one presented in Figure 3.3, where very large firms (500+) appear to experience higher constraints vis-à-vis, e.g. large ones (250-500). These dynamics seem also to emerge during years in which, overall, shares of constrained firms are higher – i.e. in 2015, 2019, 2020 – possibly indicating specific non-monotonic relationships emerging across the business cycle. Nonetheless, it must be remarked that these dynamics are descriptive, and do not account for other relevant firm characteristics which may relate to perceived constraints (see Section 4 for the econometric analysis).

indicators such as ROA and ROE are key to measuring firm profitability. They account for the return of Net Income over Assets and Equity, respectively. Profitability measures should help capture, at least to some extent, firms' operational efficiency, e.g. in managing capital. Higher levels of such indicators should thus be related, in principle, to lower financing obstacles. For the debt indicator, we use measures instead of the ratio of the firm total debt and assets, proxying firms' financial structure. Firms significantly funded by debt – i.e. with a high debt ratio – may result riskier vis-à-vis firms with lower levels of debt per unit of assets, and thus likely experience higher obstacles in accessing external finance. Finally, liquidity measures – such as the Current ratio – are aimed at capturing the ability to pay current (short-term) liabilities (debts and payables) with current assets. Hence, the higher the current ratio, the more a firm is liquid – i.e. capable of paying its obligations without external financing.

Figure 3.4 - Yearly share of firms reporting perceived credit constraints by firm productivity quantile. Years 2015-2020 (a)



Source: Authors' elaborations on BCS-BR-CCs database
(a) Labour productivity is computed as value added per employee. Values are not deflated, while productivity classes are computed, yearly, within region-sector. Regions refer to the 20 Italian regions and sector refers to the 2-digit NACE Rev. 2 classification.

Table 3.1 shows descriptive statistics for such indicators, distinguishing among firms reporting or not reporting perceived credit constraints for the period 2015-2020 (Section 2.2). Notably, it appears that firms experiencing higher obstacles in accessing external finance are those with lower levels of liquidity (Current Ratio), higher debts (over assets), and lower profitability (as proxied by ROE and ROA indicators). These relations seem to hold across all the years considered in the analysis. Interestingly, average levels for the four indicators seem to follow a similar time pattern vis-à-vis the one reported in Figure 3.1 above.

Table 3.1 - Yearly average values of financial variables for constrained and unconstrained firms (a)

Years	Perceived credit constraints	Liquidity index (mean)	Debt index (mean)	ROE index (mean)	ROA index (mean)
2015	No	1.26	0.58	4.58	4.45
2015	Yes	1.14	0.64	-4.69	2.40
2016	No	1.33	0.56	6.45	5.14
2016	Yes	1.10	0.64	0.59	2.35
2017	No	1.31	0.56	7.70	5.53
2017	Yes	1.14	0.64	0.25	2.89
2018	No	1.20	0.56	6.11	5.19
2018	Yes	0.99	0.62	-0.88	2.10
2019	No	1.23	0.55	7.20	5.56
2019	Yes	1.00	0.63	-3.12	2.97
2020	No	1.36	0.51	4.25	3.95
2020	Yes	1.18	0.55	-0.49	2.05

Source: Authors' elaborations on BCS-BR-CCs databases

(a) The variable 'perceived credit constraints' is equal to 1 (= Yes) if a firm is classified as credit-constrained (Section 2). Comparable yearly rankings across constrained and unconstrained firms are derived using median values for all variables considered. Similar results for profitability are used employing the Return On Investments (ROI) index and the Quick Ratio for liquidity.

4. Empirical strategy

In the previous subsection, we discussed a series of descriptive results about perceived firm credit constraints in Italy, over the period 2015-2020. Relying on the perceived credit constraint dummy indicator we derived in Section 2.2, results showed that the share of constrained firms tend to vary significantly across years, regions, and types of firms. Low productivity and, to some extent, small firms seem those experiencing higher obstacles in accessing external credit. In addition, constrained firms seem those with lower levels of profitability, liquidity, and higher leverage, on average. Finally, firms in Southern Italian regions appear those experiencing higher constraints vis-à-vis firms in the North and Centre of Italy (regardless of the year considered).

While this descriptive evidence may prove informative, results might depend on several other internal and external firm characteristics. For example, perceived credit constraints could be related to specific sectoral dynamics or to compositional effects along the size, productivity, and financial dimensions. It is thus useful to perform a regression analysis to account for such unobserved characteristics and control for potential compositional effects. To do that, we estimate the following linear probability equation:

$$CC_{i,t} = \alpha + \beta X_{i,t} + FE_{i,t} + \varepsilon_{i,t}$$

where $\boldsymbol{CC_{i,t}}$ is the – annual, firm-level – perceived credit constraint dummy variable (presented in Section 2). $X_{i,t}$ is a set of controls including, according to different model specifications, firm characteristics – such as size and productivity classes – as well as the financial variables described above – profitability, liquidity, and solvency indicators (Section 3). $FE_{i,t}$ captures firm-level fixed effects – such as macroregion and industry characteristics – as well as annual fixed effects.

We estimate the model for the period 2015-2020, analysing the contemporaneous relation among firm characteristics and the perceived measure of credit constraints. Additional robustness versions are estimated through Logit regressions and excluding certain years (see subsection 5.1 and Appendix A). Benchmarking our results with the evidence emerging from the literature, the analysis acts as a test regarding the reliability and usefulness of the indicator.

5. Results

In this work, we aim to investigate the relationship between a measure of perceived credit constraints, on the one hand, and firm internal and external characteristics, on the other. Table 5.1 shows the estimation results of the linear probability model (presented in the previous Section) we employ to study such relationships. Table columns progressively include the control variables described above: firm size, productivity, profitability, solvency, and liquidity, as well as industry and regional characteristics[14].

Overall, the results align with the descriptive evidence presented in Section 3 and with the hypotheses emerging from the literature. Indeed, smaller firms appear to be suffering (reporting) higher difficulties in accessing external credit. Indeed, coefficients appear negative and significant for higher-size classes (vis-à-vis the reference class), indicating a lower likelihood of perceiving constraints for larger firms[15]. In particular, when accounting for financial[16] characteristics (models 6-11), the magnitude of coefficients seems overall increasing with size classes, i.e. the higher the size class, the lower the likelihood of experiencing credit constraints[17].

The same relationship seems to hold for firms located at the bottom of the productivity distribution. In particular, coefficients for more productive firms appear negative and significant, meaning that highly productive firms have lower probabilities of appearing constrained vis-à-vis low productivity firms[18]. As discussed in Section 1, the literature finds that Small and Medium Enterprises (SMEs), as well as low-productivity firms, are usually the most affected by credit constraints: a result that seems to be confirmed here. Indeed, it may be the case that small and low-productivity firms have less transparent track records, lower collaterals, and carry out activities which are more difficult to evaluate vis-à-vis larger firms, resulting in higher costs for external credit and, likely, higher perceived obstacles in accessing external finance.

14 Table A2 and Table A3 in Appendix A include results for other specifications of our indicator.
15 Results are confirmed using the size class 50-250 as reference as well as employing the continuous (log) employment variable.
16 Regressors are overall characterised by low cross-correlations (results available upon request).
17 For the size class 500+, the magnitude is slightly lower vis-à-vis, e.g. the 250-500 class for models 1-5, indicating a non-monotonic relationship among size classes and likelihood of perceiving constraints (the same pattern observed in Figure 3.3). Nonetheless, when accounting for other firm characteristics (e.g. model 10, 11), the relationship appears non-decreasing. See also additional results in Appendix A.
18 Results are confirmed using the class 30%-70% as reference class and the continuous (log) productivity variable. See also note 13 and Table 4.

Table 5.1 - Perceived credit constraints and firm characteristics: regression results for the period 2015-2020 (a), (b)

VARIABLES	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10	Model 11
Productivity class: 10%-30%	-0.00728	-0.00894		-0.0101	-0.0111	-0.0118	-0.0117	-0.00767	-0.0100	-0.0113	-0.00989
	(0.00767)	(0.00764)		(0.00764)	(0.00763)	(0.00766)	(0.00758)	(0.00764)	(0.00765)	(0.00763)	(0.00763)
Productivity class: 30%-70%	-0.0220***	-0.0239***		-0.0278***	-0.0270***	-0.0275***	-0.0271***	-0.0211***	-0.0251***	-0.0259***	-0.0237***
	(0.00714)	(0.00712)		(0.00711)	(0.00713)	(0.00716)	(0.00708)	(0.00716)	(0.00716)	(0.00714)	(0.00716)
Productivity class: 70%-90%	-0.0270***	-0.0276***		-0.0331***	-0.0300***	-0.0301***	-0.0321***	-0.0225***	-0.0274***	-0.0300***	-0.0271***
	(0.00741)	(0.00739)		(0.00738)	(0.00739)	(0.00743)	(0.00736)	(0.00743)	(0.00743)	(0.00744)	(0.00748)
Productivity class: Top 10%	-0.0196***	-0.0249***		-0.0286***	-0.0312***	-0.0314***	-0.0347***	-0.0234***	-0.0286***	-0.0326***	-0.0294***
	(0.00741)	(0.00745)		(0.00744)	(0.00747)	(0.00751)	(0.00746)	(0.00750)	(0.00751)	(0.00754)	(0.00757)
Size class: 10-50	-0.0221***	-0.0248***	-0.0264***		-0.0235***	-0.0237***	-0.0229***	-0.0222***	-0.0230***	-0.0224***	-0.0221***
	(0.00365)	(0.00399)	(0.00397)		(0.00399)	(0.00400)	(0.00398)	(0.00399)	(0.00399)	(0.00399)	(0.00399)
Size class: 50-250	-0.0273***	-0.0368***	-0.0403***		-0.0343***	-0.0348***	-0.0359***	-0.0345***	-0.0346***	-0.0360***	-0.0357***
	(0.00400)	(0.00472)	(0.00469)		(0.00471)	(0.00473)	(0.00470)	(0.00470)	(0.00471)	(0.00471)	(0.00471)
Size class: 250-500	-0.0422***	-0.0475***	-0.0502***		-0.0435***	-0.0434***	-0.0489***	-0.0446***	-0.0444***	-0.0488***	-0.0483***
	(0.00712)	(0.00755)	(0.00753)		(0.00755)	(0.00771)	(0.00754)	(0.00754)	(0.00756)	(0.00769)	(0.00768)
Size class: 500+	-0.0359***	-0.0452***	-0.0454***		-0.0434***	-0.0432***	-0.0496***	-0.0436***	-0.0440***	-0.0492***	-0.0483***
	(0.00794)	(0.00792)	(0.00793)		(0.00791)	(0.00814)	(0.00797)	(0.00791)	(0.00791)	(0.00819)	(0.00819)
Liquidity Index						-0.00172**				0.000175	0.000212
						(0.000724)				(0.000242)	(0.000242)
Debt Index							0.0814***			0.0795***	0.0727***
							(0.00664)			(0.00676)	(0.00692)
ROA Index								-0.00129***			-0.000840***
								(0.000150)			(0.000153)
ROE Index									-0.000193***	-0.000163***	
									(3.75e-05)	(3.76e-05)	
Observations	30,579	30,579	30,579	30,579	30,579	30,395	30,579	30,579	30,579	30,395	30,395
R-squared	0.009	0.025	0.024	0.022	0.027	0.027	0.033	0.029	0.028	0.034	0.034
Industry Fixed Effects	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region Fixed Effects	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Source: Authors' elaborations on BCS-BR-CCs databases

(a) The Table shows results for the linear probability model described in Section 4, where the dependent variable is a dummy equal to 1 if the firm perceives credit constraints each year and 0 otherwise (Section 2.2).
The Table does not report year and macroregion coefficients. The complete list of coefficients is available in Table A1 in Appendix A. Missing size class and productivity class coefficients are not reported. Productivity classes are computed within Region-Sector. Regions refer to the 20 Italian regions while sectors refer to the 2-dig NACE Rev. 2 sectoral classification. Reference categories are Year = 2015, Size Class 1-9, and Productivity class 10%. Results are robust using the productivity class 30%-70% and size class 50-250 as reference classes and using the continuous log(size) and log(productivity) variables. Results are robust also using industry-region fixed effects (Table A4), excluding the years 2015 and 2020 (Table A5), and employing a Logit model (Table A6). Robust standard errors are in parentheses.

(b) *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Focussing on financial characteristics, firms with higher liquidity (model 6) and higher profitability (model 8, 9) are those experiencing lower constraints (with the other coefficients remaining significant). Conversely, firms with higher levels of debt (model 7) are associated to higher perceived credit constraints. This may be because firms with higher liquidity can use internal resources more easily in the short run, not having to rely on external finance to fund their operations. Similarly, firms with higher profitability (ROA, ROE indicators) may instead be perceived as less risky by credit intermediaries, thus having lower obstacles (or costs) in accessing external credit[19]. On the contrary, firms that are more indebted may present higher risks, likely related to lower ability of sustaining the cost of debt and constraining access to external finance. However, if one simultaneously controls for all the above-mentioned financial characteristics (models 10, 11), the liquidity coefficient becomes not significant while solvency and profitability indicators do remain significant. This may point to the higher relevance of long-term measures (such as debts, equity and assets) vis-à-vis short-term ones for assessing credit risk.

The relationships commented on above hold controlling for year[20], industry and regional fixed effects (see the complete Table A 1 in Appendix A). Notably, taking the North-West NUTS-1 macroregion as a reference class, firms located in the Italian Mezzogiorno (South and Islands) appear more likely to perceive obstacles in accessing external credit (positive and significant coefficients). On the contrary, firms in the Centre and the North-East do not experience significantly lower constraints vis-à-vis the reference class. This figure may be related to the overall lower development of the credit and financial system in the Mezzogiorno[21].

5.1 Robustness checks, limitations, and future research

To ensure the robustness of our findings, we conducted several robustness checks. First, we tested alternative specifications of the perceived credit constraint indicator to perform a sensitivity analysis of the results. As explained in Section 2.2, we derive our primary perceived credit constraint indicator starting from quarterly data at the firm level. We categorise firms as credit-constrained in a given year if they report perceived obstacles in accessing external finance in more than half of the available quarters. To

19 This result holds also using the ROI indicator (results are available upon request).

20 Relative to 2015, year coefficients seem to capture the overall descriptive dynamics observed in Section 3 (Figure 3.1). Results are robust excluding the first (2015) and final (2020) years from the sample, i.e. focussing the analysis on the period 2016-2019 (Table A5).

21 The results are robust also controlling for Industry-Region fixed effects (Appendix A).

further test the robustness of our results, we derive two additional measures. Firstly, we categorise firms as credit-constrained if they report obstacles in accessing external credit for all quarters. Secondly, we categorise firms as credit-constrained if they have more than three-quarters available and report perceived obstacles more than half of the time[22]. As shown in Table A2 and Table A3 in Appendix A, the main findings are confirmed also using these additional specifications. Indeed, as commented in Section 5, smaller, less productive, and financially unhealthier firms result in those experiencing higher perceived constraints.

To evaluate the stability of the results, we also employed two additional model specifications, which again confirm the relationships described above. Employing a Logit model and deriving the marginal effects, we find the same significance levels and magnitudes for all relevant coefficients (Table A6 in Appendix A). We also find comparable results by specifying the usual linear probability model using the period 2016-2019 (Table A5). On the one hand, the Logit model allows testing the stability of results in light of potential limitations of linear probability models (e.g. predicted probabilities that fall outside the valid probability range of 0 to 1). On the other hand, the exclusion of the first (2015) and last year (2020) ensures that results are not driven by the pandemic year and/or by the less robust perceived credit constraint indicator derived for 2015 (for which only the fourth quarter is available - see Section 2.2).

Overall, the analysis proves robust to a series of robustness checks. However, it is important to remark that statistically significant relationships between variables are limited to correlations and do not establish causal relationships. While this study incorporates several firm controls to mitigate potential endogeneity concerns, the observed relationship between financial conditions and perceived credit constraints likely requires further investigation to avoid potential reverse causality issues. Future research may apply an IV approach and the increased availability of data on firms over multiple years to further assess those relationships. Despite these limitations, results are in line with the existing literature and reassure about the reliability of the derived perceived credit constraint indicator.

22 Hence, firms with three quarters available are classified as constrained if they report perceived obstacles in two or three of the quarters. Firms with four quarters available are classified as constrained if they report obstacles in in three or four of the quarters.

6. Conclusions

This work has analysed the relationship between credit constraints and firm characteristics in Italy. We have used qualitative firm-level quarterly information from the Istat's *Business Confidence Survey*, deriving an annual measure of perceived credit constraints for 2015-2020. Using a linear probability model, we have tested the indicator's informative power by relating it to firm financials and other firm characteristics from the Italian Business Register. In line with the existing literature, the analysis has highlighted that constraints seem more binding for smaller, less productive firms in Southern Italian regions. Relatedly, we have explored the relationship between the indicator and financial conditions at the firm level, finding that financially healthier firms experience lower credit access obstacles. In particular, firms with higher profitability seem less affected by credit constraints, while firms with higher levels of debt experience higher perceived obstacles in accessing external credit. The analysis brings novel empirical evidence about perceived measures of credit constraints in Italy over several years and provides a potentially useful indicator for firm-level empirical and policy analyses. In future research, the perceived measure of credit constraints can be used to examine how public policies influence credit access and perceived constraints and to investigate business dynamics over the COVID-19 pandemic.

Appendix A

Table A1 - Perceived credit constraints and firm characteristics: regression results for the period 2015-2020 (complete table) (a), (b)

VARIABLES	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10	Model 11
Year: 2016	-0.0323***	-0.0330***	-0.0322***	-0.0360***	-0.0336***	-0.0333***	-0.0316***	-0.0324***	-0.0330***	-0.0309***	-0.0308***
	(0.00550)	(0.00545)	(0.00546)	(0.00543)	(0.00545)	(0.00547)	(0.00543)	(0.00544)	(0.00544)	(0.00544)	(0.00545)
Year: 2017	-0.0571***	-0.0583***	-0.0582***	-0.0595***	-0.0585***	-0.0584***	-0.0568***	-0.0570***	-0.0578***	-0.0563***	-0.0560***
	(0.00548)	(0.00545)	(0.00545)	(0.00547)	(0.00544)	(0.00544)	(0.00542)	(0.00544)	(0.00544)	(0.00542)	(0.00542)
Year: 2018	-0.0507***	-0.0516***	-0.0513***	-0.0525***	-0.0527***	-0.0528***	-0.0508***	-0.0514***	-0.0522***	-0.0505***	-0.0503***
	(0.00547)	(0.00545)	(0.00545)	(0.00547)	(0.00545)	(0.00545)	(0.00543)	(0.00545)	(0.00544)	(0.00543)	(0.00543)
Year: 2019	-0.0343***	-0.0353***	-0.0350***	-0.0364***	-0.0362***	-0.0363***	-0.0340***	-0.0345***	-0.0356***	-0.0336***	-0.0332***
	(0.00573)	(0.00570)	(0.00570)	(0.00572)	(0.00570)	(0.00570)	(0.00568)	(0.00569)	(0.00569)	(0.00567)	(0.00568)
Year: 2020	-0.0462***	-0.0471***	-0.0468***	-0.0489***	-0.0473***	-0.0471***	-0.0415***	-0.0477***	-0.0472***	-0.0416***	-0.0425***
	(0.00563)	(0.00561)	(0.00561)	(0.00564)	(0.00561)	(0.00561)	(0.00560)	(0.00561)	(0.00561)	(0.00559)	(0.00560)
Productivity class: 10%-30%	-0.00728	-0.00894		-0.0101	-0.0111	-0.0118	-0.0117	-0.00767	-0.0100	-0.0113	-0.00989
	(0.00767)	(0.00764)		(0.00764)	(0.00763)	(0.00766)	(0.00758)	(0.00764)	(0.00765)	(0.00763)	(0.00763)
Productivity class: 30%-70%	-0.0220***	-0.0239***		-0.0278***	-0.0270***	-0.0275***	-0.0271***	-0.0211***	-0.0251***	-0.0259***	-0.0237***
	(0.00714)	(0.00712)		(0.00711)	(0.00713)	(0.00716)	(0.00708)	(0.00716)	(0.00716)	(0.00714)	(0.00716)
Productivity class: 70%-90%	-0.0270***	-0.0276***		-0.0331***	-0.0300***	-0.0301***	-0.0321***	-0.0225***	-0.0274***	-0.0300***	-0.0271***
	(0.00741)	(0.00739)		(0.00738)	(0.00739)	(0.00743)	(0.00736)	(0.00743)	(0.00743)	(0.00744)	(0.00748)
Productivity class: Top 10%	-0.0196***	-0.0249***		-0.0286***	-0.0312***	-0.0314***	-0.0347***	-0.0234***	-0.0286***	-0.0326***	-0.0294***
	(0.00741)	(0.00745)		(0.00744)	(0.00747)	(0.00751)	(0.00746)	(0.00750)	(0.00751)	(0.00754)	(0.00757)
Size class: 10-50	-0.0221***	-0.0248***	-0.0264***		-0.0235***	-0.0237***	-0.0229***	-0.0222***	-0.0230***	-0.0224***	-0.0221***
	(0.00365)	(0.00399)	(0.00397)		(0.00399)	(0.00400)	(0.00398)	(0.00399)	(0.00399)	(0.00399)	(0.00399)
Size class: 50-250	-0.0273***	-0.0368***	-0.0403***		-0.0343***	-0.0348***	-0.0359***	-0.0345***	-0.0346***	-0.0360***	-0.0357***
	(0.00400)	(0.00472)	(0.00469)		(0.00471)	(0.00473)	(0.00470)	(0.00470)	(0.00471)	(0.00471)	(0.00471)
Size class: 250-500	-0.0422***	-0.0475***	-0.0502***		-0.0435***	-0.0434***	-0.0489***	-0.0446***	-0.0444***	-0.0488***	-0.0483***
	(0.00712)	(0.00755)	(0.00753)		(0.00755)	(0.00771)	(0.00754)	(0.00754)	(0.00756)	(0.00769)	(0.00768)
Size class: 500+	-0.0359***	-0.0452***	-0.0454***		-0.0434***	-0.0432***	-0.0496***	-0.0436***	-0.0440***	-0.0492***	-0.0483***
	(0.00794)	(0.00792)	(0.00793)		(0.00791)	(0.00814)	(0.00797)	(0.00791)	(0.00791)	(0.00819)	(0.00819)
NUTS1: North-East					0.000571	0.000575	-8.61e-05	0.000603	0.000730	0.000193	0.000130
					(0.00329)	(0.00331)	(0.00329)	(0.00329)	(0.00329)	(0.00330)	(0.00330)
NUTS1: Centre					0.00562	0.00578	0.00307	0.00540	0.00577	0.00355	0.00350
					(0.00366)	(0.00368)	(0.00366)	(0.00366)	(0.00366)	(0.00367)	(0.00367)
NUTS1: South					0.0313***	0.0315***	0.0282***	0.0297***	0.0309***	0.0284***	0.0279***
					(0.00497)	(0.00499)	(0.00495)	(0.00495)	(0.00496)	(0.00497)	(0.00497)

Table A1 cont- - Perceived credit constraints and firm characteristics: regression results for the period 2015-2020 (complete table) (a), (b)

VARIABLES	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10	Model 11
NUTS1: Islands					0.0274***	0.0276***	0.0250***	0.0249***	0.0270***	0.0249***	0.0238***
					(0.00808)	(0.00809)	(0.00802)	(0.00805)	(0.00805)	(0.00801)	(0.00802)
Liquidity Index						-0.00172**				0.000175	0.000212
						(0.000724)				(0.000242)	(0.000242)
Debt Index							0.0814***			0.0795***	0.0727***
							(0.00664)			(0.00676)	(0.00692)
ROA Index								-0.00129***			-0.000840***
								(0.000150)			(0.000153)
ROE Index									-0.000193***	-0.000163***	
									(3.75e-05)	(3.76e-05)	
Observations	30,579	30,579	30,579	30,579	30,579	30,395	30,579	30,579	30,579	30,395	30,395
R-squared	0.009	0.025	0.024	0.022	0.027	0.027	0.033	0.029	0.028	0.034	0.034
Industry Fixed Effects	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region Fixed Effects	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
DF	15	15	10	10	19	20	20	20	20	22	22
Adj. R2	0.00852	0.0219	0.0210	0.0192	0.0238	0.0242	0.0298	0.0264	0.0250	0.0307	0.0309

Source: Authors' elaborations on BCS-BR-CCs databases

(a) The table shows results for the linear probability model described in Section 4, where the dependent variable is a dummy equal to 1 if the firm perceives credit constraints each year and 0 otherwise (Section 2.2). Missing size class and productivity class coefficients are not reported. Productivity classes are computed within Region-Sector. Regions refer to the 20 Italian regions while sectors refer to the 2-dig NACE Rev. 2 sectoral classification. Reference categories are Year = 2015, Size Class 1-9, and Productivity class 10%. Results are robust using the productivity class 30%-70% and size class 50-250 as reference classes and using the continuous log(size) and log(productivity) variables. Results are robust also using industry-region fixed effects (Table A4), excluding the years 2015 and 2020 (Table A5), and employing a Logit model (Table A6). Robust standard errors are in parentheses.

(b) *** p<0.01, ** p<0.05, * p<0.1

Table A2 - Alternative perceived credit constraint indicator (A) and firm characteristics: regression results (a), (b)

VARIABLES	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10	Model 11
Year: 2016	-0.0499***	-0.0507***	-0.0500***	-0.0528***	-0.0512***	-0.0505***	-0.0496***	-0.0503***	-0.0507***	-0.0486***	-0.0486***
	(0.00529)	(0.00525)	(0.00526)	(0.00524)	(0.00525)	(0.00527)	(0.00523)	(0.00525)	(0.00524)	(0.00525)	(0.00525)
Year: 2017	-0.0757***	-0.0769***	-0.0768***	-0.0777***	-0.0770***	-0.0770***	-0.0758***	-0.0760***	-0.0765***	-0.0754***	-0.0753***
	(0.00511)	(0.00510)	(0.00510)	(0.00512)	(0.00509)	(0.00509)	(0.00507)	(0.00509)	(0.00509)	(0.00507)	(0.00507)
Year: 2018	-0.0709***	-0.0719***	-0.0715***	-0.0725***	-0.0727***	-0.0728***	-0.0714***	-0.0719***	-0.0724***	-0.0711***	-0.0709***
	(0.00514)	(0.00513)	(0.00513)	(0.00515)	(0.00513)	(0.00514)	(0.00511)	(0.00513)	(0.00513)	(0.00511)	(0.00511)
Year: 2019	-0.0635***	-0.0645***	-0.0643***	-0.0653***	-0.0653***	-0.0653***	-0.0637***	-0.0641***	-0.0648***	-0.0633***	-0.0631***
	(0.00526)	(0.00525)	(0.00525)	(0.00527)	(0.00525)	(0.00525)	(0.00523)	(0.00525)	(0.00525)	(0.00523)	(0.00523)
Year: 2020	-0.0731***	-0.0741***	-0.0738***	-0.0754***	-0.0743***	-0.0742***	-0.0700***	-0.0746***	-0.0742***	-0.0701***	-0.0706***
	(0.00513)	(0.00512)	(0.00512)	(0.00514)	(0.00512)	(0.00512)	(0.00509)	(0.00511)	(0.00511)	(0.00509)	(0.00509)
Productivity class: 10%-30%	-0.0104	-0.0119*		-0.0128*	-0.0136**	-0.0139**	-0.0140**	-0.0112*	-0.0128*	-0.0135**	-0.0127*
	(0.00668)	(0.00667)		(0.00666)	(0.00667)	(0.00669)	(0.00663)	(0.00668)	(0.00668)	(0.00668)	(0.00668)
Productivity class: 30%-70%	-0.0232***	-0.0250***		-0.0277***	-0.0274***	-0.0275***	-0.0275***	-0.0233***	-0.0260***	-0.0263***	-0.0249***
	(0.00624)	(0.00625)		(0.00623)	(0.00626)	(0.00628)	(0.00622)	(0.00628)	(0.00628)	(0.00628)	(0.00629)
Productivity class: 70%-90%	-0.0245***	-0.0254***		-0.0293***	-0.0273***	-0.0272***	-0.0288***	-0.0220***	-0.0253***	-0.0271***	-0.0253***
	(0.00647)	(0.00646)		(0.00645)	(0.00647)	(0.00650)	(0.00645)	(0.00649)	(0.00650)	(0.00652)	(0.00654)
Productivity class: Top 10%	-0.0199***	-0.0244***		-0.0270***	-0.0293***	-0.0294***	-0.0319***	-0.0239***	-0.0273***	-0.0302***	-0.0282***
	(0.00647)	(0.00652)		(0.00649)	(0.00654)	(0.00657)	(0.00654)	(0.00657)	(0.00658)	(0.00661)	(0.00663)
Size class: 10-50	-0.0168***	-0.0177***	-0.0192***		-0.0167***	-0.0168***	-0.0163***	-0.0158***	-0.0163***	-0.0159***	-0.0157***
	(0.00306)	(0.00337)	(0.00336)		(0.00336)	(0.00337)	(0.00336)	(0.00337)	(0.00336)	(0.00336)	(0.00337)
Size class: 50-250	-0.0214***	-0.0264***	-0.0294***		-0.0245***	-0.0247***	-0.0257***	-0.0246***	-0.0247***	-0.0256***	-0.0254***
	(0.00332)	(0.00393)	(0.00391)		(0.00392)	(0.00394)	(0.00392)	(0.00391)	(0.00392)	(0.00393)	(0.00393)
Size class: 250-500	-0.0293***	-0.0319***	-0.0339***		-0.0287***	-0.0280***	-0.0327***	-0.0295***	-0.0294***	-0.0321***	-0.0317***
	(0.00610)	(0.00648)	(0.00646)		(0.00648)	(0.00668)	(0.00648)	(0.00648)	(0.00649)	(0.00667)	(0.00666)
Size class: 500+	-0.0272***	-0.0340***	-0.0341***		-0.0326***	-0.0316***	-0.0372***	-0.0328***	-0.0331***	-0.0362***	-0.0355***
	(0.00652)	(0.00649)	(0.00650)		(0.00648)	(0.00676)	(0.00653)	(0.00648)	(0.00649)	(0.00681)	(0.00680)
NUTS1: North-East					0.00219	0.00196	0.00171	0.00222	0.00231	0.00168	0.00162
					(0.00272)	(0.00273)	(0.00271)	(0.00272)	(0.00272)	(0.00273)	(0.00273)
NUTS1: Centre					0.00503*	0.00489	0.00315	0.00488	0.00515*	0.00322	0.00316
					(0.00301)	(0.00303)	(0.00301)	(0.00301)	(0.00301)	(0.00303)	(0.00303)
NUTS1: South					0.0250***	0.0249***	0.0227***	0.0239***	0.0247***	0.0225***	0.0223***
					(0.00417)	(0.00419)	(0.00416)	(0.00416)	(0.00417)	(0.00418)	(0.00418)

Table A2 cont. - Alternative perceived credit constraint indicator (A) and firm characteristics: regression results (a), (b)

VARIABLES	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10	Model 11
NUTS1: Islands					0.0213***	0.0212***	0.0195***	0.0195***	0.0210***	0.0192***	0.0184***
					(0.00677)	(0.00679)	(0.00673)	(0.00676)	(0.00675)	(0.00672)	(0.00674)
Liquidity Index						-0.00103**				0.000403	0.000425
						(0.000448)				(0.000258)	(0.000267)
Debt Index							0.0601***			0.0597***	0.0553***
							(0.00560)			(0.00573)	(0.00582)
ROA Index								-0.000897***			-0.000569***
								(0.000124)			(0.000126)
ROE Index									-0.000147***	-0.000127***	
									(3.27e-05)	(3.28e-05)	
Observations	30,579	30,579	30,579	30,579	30,579	30,395	30,579	30,579	30,579	30,395	30,395
R-squared	0.018	0.029	0.028	0.027	0.031	0.031	0.036	0.033	0.032	0.037	0.037
Industry Fixed Effects	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region Fixed Effects	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Source: Authors' elaborations on BCS-BR-CCs databases

(a) The table shows results for the linear probability model described in Section 4, where the dependent variable is a dummy equal to 1 if the firm perceives credit constraints each year and 0 otherwise. The indicator is in this case computed as equal to 1 if firms report obstacles in accessing external credit for all quarters (see Section 2.2 and 5.1 for additional details on the baseline indicator and additional versions). Missing size class and productivity class coefficients are not reported. Productivity classes are computed within Region-Sector. Regions refer to the 20 Italian regions while sectors refer to the 2-dig NACE Rev. 2 sectoral classification. Reference categories are Year = 2015, Size Class 1-9, and Productivity class 10%. Results are robust using the productivity class 30%-70% and size class 50-250 as reference classes and using the continuous log(size) and log(productivity) variables. Robust standard errors are in parentheses.

(b) *** p<0.01, ** p<0.05, * p<0.1.

Table A3 - Alternative perceived credit constraint indicator (B) and firm characteristics: regression results (a), (b)

VARIABLES	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10	Model 11
Year: 2017	-0.01000*	-0.0107*	-0.0106*	-0.00881	-0.0109*	-0.0111**	-0.0123**	-0.0110**	-0.0108*	-0.0119**	-0.0118**
	(0.00557)	(0.00556)	(0.00557)	(0.00543)	(0.00556)	(0.00558)	(0.00555)	(0.00555)	(0.00556)	(0.00558)	(0.00557)
Year: 2018	-0.0107**	-0.0105*	-0.0104*	-0.00858	-0.0116**	-0.0122**	-0.0127**	-0.0117**	-0.0118**	-0.0127**	-0.0125**
	(0.00543)	(0.00548)	(0.00549)	(0.00534)	(0.00549)	(0.00552)	(0.00549)	(0.00548)	(0.00549)	(0.00551)	(0.00551)
Year: 2019	0.00576	0.00585	0.00603	0.00753	0.00499	0.00451	0.00397	0.00512	0.00489	0.00410	0.00439
	(0.00584)	(0.00589)	(0.00589)	(0.00575)	(0.00588)	(0.00590)	(0.00587)	(0.00587)	(0.00587)	(0.00589)	(0.00589)
Year: 2020	-0.00373	-0.00379	-0.00346	-0.00240	-0.00421	-0.00363	-0.00261	-0.00563	-0.00463	-0.00275	-0.00352
	(0.00577)	(0.00577)	(0.00577)	(0.00563)	(0.00578)	(0.00580)	(0.00574)	(0.00578)	(0.00577)	(0.00576)	(0.00577)
Productivity class: 10%-30%	-0.00788	-0.00895		-0.00829	-0.0110	-0.0117	-0.0131	-0.00916	-0.0101	-0.0129	-0.0122
	(0.0115)	(0.0115)		(0.0114)	(0.0115)	(0.0115)	(0.0114)	(0.0115)	(0.0115)	(0.0114)	(0.0115)
Productivity class: 30%-70%	-0.0267**	-0.0278***		-0.0285***	-0.0310***	-0.0314***	-0.0327***	-0.0271**	-0.0295***	-0.0319***	-0.0304***
	(0.0108)	(0.0108)		(0.0107)	(0.0108)	(0.0108)	(0.0108)	(0.0108)	(0.0108)	(0.0108)	(0.0109)
Productivity class: 70%-90%	-0.0345***	-0.0352***		-0.0366***	-0.0382***	-0.0384***	-0.0412***	-0.0330***	-0.0361***	-0.0397***	-0.0376***
	(0.0110)	(0.0110)		(0.0109)	(0.0110)	(0.0110)	(0.0110)	(0.0111)	(0.0110)	(0.0110)	(0.0111)
Productivity class: Top 10%	-0.0202*	-0.0236**		-0.0241**	-0.0299***	-0.0301***	-0.0339***	-0.0246**	-0.0278**	-0.0322***	-0.0300***
	(0.0111)	(0.0111)		(0.0111)	(0.0112)	(0.0112)	(0.0112)	(0.0112)	(0.0112)	(0.0112)	(0.0113)
Size class: 10-50	-0.0119**	-0.0163***	-0.0178***		-0.0155***	-0.0167***	-0.0153***	-0.0146***	-0.0153***	-0.0154***	-0.0150***
	(0.00462)	(0.00522)	(0.00521)		(0.00522)	(0.00525)	(0.00522)	(0.00522)	(0.00521)	(0.00524)	(0.00525)
Size class: 50-250	-0.0112**	-0.0211***	-0.0247***		-0.0193***	-0.0211***	-0.0212***	-0.0194***	-0.0196***	-0.0213***	-0.0209***
	(0.00528)	(0.00628)	(0.00625)		(0.00626)	(0.00630)	(0.00625)	(0.00625)	(0.00626)	(0.00630)	(0.00629)
Size class: 250-500	-0.0242***	-0.0308***	-0.0331***		-0.0281***	-0.0317***	-0.0326***	-0.0291***	-0.0287***	-0.0339***	-0.0336***
	(0.00884)	(0.00953)	(0.00946)		(0.00951)	(0.00963)	(0.00951)	(0.00953)	(0.00953)	(0.00962)	(0.00961)
Size class: 500+	-0.0170*	-0.0292***	-0.0284***		-0.0288***	-0.0324***	-0.0348***	-0.0287***	-0.0292***	-0.0357***	-0.0347***
	(0.00989)	(0.00933)	(0.00934)		(0.00928)	(0.00927)	(0.00943)	(0.00928)	(0.00926)	(0.00934)	(0.00936)
NUTS1: North-East					-0.00577	-0.00500	-0.00601	-0.00578	-0.00574	-0.00535	-0.00537
					(0.00414)	(0.00415)	(0.00413)	(0.00414)	(0.00414)	(0.00414)	(0.00414)
NUTS1: Centre					-0.000618	0.000174	-0.00171	-0.000794	-0.000625	-0.000902	-0.000920
					(0.00470)	(0.00471)	(0.00469)	(0.00470)	(0.00470)	(0.00470)	(0.00471)
NUTS1: South					0.0267***	0.0270***	0.0246***	0.0254***	0.0263***	0.0251***	0.0247***
					(0.00669)	(0.00671)	(0.00667)	(0.00668)	(0.00669)	(0.00669)	(0.00669)

Table A3 cont. - Alternative perceived credit constraint indicator (B) and firm characteristics: regression results (a), (b)

VARIABLES	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10	Model 11
NUTS1: Islands					0.0238**	0.0245**	0.0222**	0.0217**	0.0234**	0.0224**	0.0215**
					(0.0110)	(0.0110)	(0.0109)	(0.0109)	(0.0109)	(0.0109)	(0.0109)
Liquidity Index						-0.00563***				-0.00122	-0.00105
						(0.00118)				(0.00126)	(0.00126)
Debt Index							0.0558***			0.0504***	0.0453***
							(0.00876)			(0.00975)	(0.00994)
ROA Index								-0.000992***			-0.000671***
								(0.000186)			(0.000189)
ROE Index									-0.000147***	-0.000122**	
									(4.85e-05)	(4.85e-05)	
Observations	13,863	13,863	13,863	13,863	13,863	13,786	13,863	13,863	13,863	13,786	13,786
R-squared	0.004	0.020	0.018	0.018	0.022	0.024	0.026	0.024	0.023	0.027	0.027
Industry Fixed Effects	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region Fixed Effects	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Source: Authors' elaborations on BCS-BR-CCs databases

(a) The table shows results for the linear probability model described in Section 4, where the dependent variable is a dummy equal to 1 if the firm perceives credit constraints each year and 0 otherwise. The indicator is in this case computed as equal to 1 if firms report obstacles in accessing external credit for more than half of the times and have more than three quarters available each year (see Section 2.2 and 5.1 for additional details on the baseline indicator and additional versions). Missing size class and productivity class coefficients are not reported. Productivity classes are computed within Region-Sector. Regions refer to the 20 Italian regions while sectors refer to the 2-dig NACE Rev. 2 sectoral classification. Reference categories are Year = 2015, Size Class 1-9, and Productivity class 10%. Results are robust using the productivity class 30%-70% and size class 50-250 as reference classes and using the continuous log(size) and log(productivity) variables. Robust standard errors are in parentheses.

(b) *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table A4 - Perceived credit constraints and firm characteristics: regression results for the period 2015-2020 (additional fixed effects results) (a), (b)

VARIABLES	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 11
Year: 2016	-0.0323***	-0.0330***	-0.0323***	-0.0366***	-0.0333***	-0.0330***	-0.0312***	-0.0321***	-0.0305***
	(0.00550)	(0.00545)	(0.00536)	(0.00532)	(0.00535)	(0.00537)	(0.00533)	(0.00535)	(0.00535)
Year: 2017	-0.0571***	-0.0583***	-0.0579***	-0.0591***	-0.0581***	-0.0580***	-0.0564***	-0.0566***	-0.0557***
	(0.00548)	(0.00545)	(0.00536)	(0.00538)	(0.00536)	(0.00536)	(0.00534)	(0.00535)	(0.00534)
Year: 2018	-0.0507***	-0.0516***	-0.0521***	-0.0535***	-0.0526***	-0.0527***	-0.0507***	-0.0514***	-0.0502***
	(0.00547)	(0.00545)	(0.00538)	(0.00540)	(0.00538)	(0.00539)	(0.00536)	(0.00538)	(0.00537)
Year: 2019	-0.0343***	-0.0353***	-0.0350***	-0.0366***	-0.0354***	-0.0355***	-0.0331***	-0.0339***	-0.0325***
	(0.00573)	(0.00570)	(0.00563)	(0.00565)	(0.00563)	(0.00563)	(0.00561)	(0.00563)	(0.00561)
Year: 2020	-0.0462***	-0.0471***	-0.0459***	-0.0480***	-0.0462***	-0.0460***	-0.0403***	-0.0467***	-0.0412***
	(0.00563)	(0.00561)	(0.00555)	(0.00557)	(0.00555)	(0.00555)	(0.00554)	(0.00555)	(0.00554)
Productivity class: 10%-30%	-0.00728	-0.00894		-0.0124	-0.0111	-0.0117	-0.0119	-0.00786	-0.0102
	(0.00767)	(0.00764)		(0.00766)	(0.00766)	(0.00769)	(0.00762)	(0.00767)	(0.00767)
Productivity class: 30%-70%	-0.0220***	-0.0239***		-0.0306***	-0.0269***	-0.0273***	-0.0272***	-0.0211***	-0.0239***
	(0.00714)	(0.00712)		(0.00716)	(0.00716)	(0.00719)	(0.00712)	(0.00719)	(0.00720)
Productivity class: 70%-90%	-0.0270***	-0.0276***		-0.0354***	-0.0303***	-0.0304***	-0.0326***	-0.0230***	-0.0279***
	(0.00741)	(0.00739)		(0.00744)	(0.00744)	(0.00748)	(0.00741)	(0.00749)	(0.00753)
Productivity class: Top 10%	-0.0196***	-0.0249***		-0.0358***	-0.0316***	-0.0315***	-0.0359***	-0.0235***	-0.0303***
	(0.00741)	(0.00745)		(0.00757)	(0.00759)	(0.00762)	(0.00759)	(0.00763)	(0.00772)
Size class: 10-50	-0.0221***	-0.0248***	-0.0255***		-0.0233***	-0.0235***	-0.0227***	-0.0222***	-0.0220***
	(0.00365)	(0.00399)	(0.00419)		(0.00421)	(0.00422)	(0.00420)	(0.00421)	(0.00420)
Size class: 50-250	-0.0273***	-0.0368***	-0.0379***		-0.0331***	-0.0336***	-0.0348***	-0.0334***	-0.0347***
	(0.00400)	(0.00472)	(0.00493)		(0.00497)	(0.00499)	(0.00496)	(0.00496)	(0.00497)
Size class: 250-500	-0.0422***	-0.0475***	-0.0457***		-0.0419***	-0.0420***	-0.0475***	-0.0432***	-0.0473***
	(0.00712)	(0.00755)	(0.00785)		(0.00787)	(0.00803)	(0.00787)	(0.00786)	(0.00802)
Size class: 500+	-0.0359***	-0.0452***	-0.0574***		-0.0568***	-0.0576***	-0.0624***	-0.0571***	-0.0621***
	(0.00794)	(0.00792)	(0.00854)		(0.00853)	(0.00876)	(0.00858)	(0.00853)	(0.00880)
Liquidity Index						-0.00191**			-4.18e-09
						(0.000769)			(0.000275)
Debt Index							0.0847***		0.0755***
							(0.00705)		(0.00735)
ROA Index								-0.00125***	-0.000771***
								(0.000157)	(0.000161)

Table A4 cont. - Perceived credit constraints and firm characteristics: regression results for the period 2015-2020 (additional fixed effects results) (a), (b)

VARIABLES	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 11
Observations	30,579	30,579	30,579	30,579	30,579	30,395	30,579	30,579	30,395
R-squared	0.009	0.025	0.087	0.086	0.088	0.090	0.094	0.091	0.096
Industry Fixed Effects	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Source: Authors' elaborations on BCS-BR-CCs databases

(a) The table shows results for the linear probability model described in Section 4, where the dependent variable is a dummy equal to 1 if the firm perceives credit constraints each year and 0 otherwise (Section 2.2). Missing size class and productivity class coefficients are not reported. Productivity classes are computed within Region-Sector. Regions refer to the 20 Italian regions while sectors refer to the 2-dig NACE Rev. 2 sectoral classification. Reference categories are Year = 2015, Size Class 1-9, and Productivity class 10%. Robust standard errors are in parentheses.

(b) *** p<0.01, ** p<0.05, * p<0.1.

Table A5 - Perceived credit constraints and firm characteristics: regression results for the period 2016-2019 (a),(b)

VARIABLES	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10	Model 11
Year: 2017	-0.0252***	-0.0255***	-0.0262***	-0.0234***	-0.0249***	-0.0251***	-0.0253***	-0.0246***	-0.0248***	-0.0253***	-0.0251***
	(0.00417)	(0.00416)	(0.00417)	(0.00408)	(0.00415)	(0.00417)	(0.00416)	(0.00415)	(0.00415)	(0.00416)	(0.00416)
Year: 2018	-0.0187***	-0.0189***	-0.0194***	-0.0165***	-0.0192***	-0.0196***	-0.0194***	-0.0192***	-0.0193***	-0.0196***	-0.0195***
	(0.00416)	(0.00414)	(0.00416)	(0.00407)	(0.00414)	(0.00417)	(0.00416)	(0.00414)	(0.00414)	(0.00416)	(0.00415)
Year: 2019	-0.00230	-0.00248	-0.00307	-0.000364	-0.00269	-0.00307	-0.00262	-0.00229	-0.00270	-0.00269	-0.00242
	(0.00451)	(0.00449)	(0.00450)	(0.00441)	(0.00449)	(0.00451)	(0.00449)	(0.00448)	(0.00448)	(0.00448)	(0.00449)
Productivity class: 10%-30%	-0.0129	-0.0146		-0.0150*	-0.0164*	-0.0174*	-0.0172*	-0.0132	-0.0154*	-0.0164*	-0.0151*
	(0.00891)	(0.00887)		(0.00888)	(0.00887)	(0.00892)	(0.00886)	(0.00887)	(0.00890)	(0.00888)	(0.00887)
Productivity class: 30%-70%	-0.0276***	-0.0299***		-0.0326***	-0.0327***	-0.0335***	-0.0328***	-0.0271***	-0.0309***	-0.0314***	-0.0292***
	(0.00832)	(0.00831)		(0.00830)	(0.00832)	(0.00837)	(0.00830)	(0.00833)	(0.00836)	(0.00834)	(0.00834)
Productivity class: 70%-90%	-0.0308***	-0.0319***		-0.0362***	-0.0341***	-0.0344***	-0.0359***	-0.0270***	-0.0316***	-0.0338***	-0.0310***
	(0.00863)	(0.00860)		(0.00861)	(0.00861)	(0.00867)	(0.00863)	(0.00863)	(0.00866)	(0.00868)	(0.00870)
Productivity class: Top 10%	-0.0256***	-0.0315***		-0.0343***	-0.0372***	-0.0376***	-0.0404***	-0.0300***	-0.0349***	-0.0384***	-0.0353***
	(0.00862)	(0.00868)		(0.00867)	(0.00869)	(0.00875)	(0.00873)	(0.00869)	(0.00873)	(0.00877)	(0.00878)
Size class: 10-50	-0.0178***	-0.0191***	-0.0208***		-0.0178***	-0.0180***	-0.0172***	-0.0165***	-0.0174***	-0.0169***	-0.0164***
	(0.00413)	(0.00450)	(0.00448)		(0.00450)	(0.00452)	(0.00450)	(0.00449)	(0.00450)	(0.00449)	(0.00449)
Size class: 50-250	-0.0231***	-0.0300***	-0.0337***		-0.0275***	-0.0278***	-0.0286***	-0.0275***	-0.0277***	-0.0289***	-0.0286***
	(0.00455)	(0.00537)	(0.00534)		(0.00535)	(0.00539)	(0.00536)	(0.00534)	(0.00535)	(0.00536)	(0.00536)
Size class: 250-500	-0.0409***	-0.0435***	-0.0465***		-0.0392***	-0.0390***	-0.0420***	-0.0399***	-0.0400***	-0.0426***	-0.0420***
	(0.00777)	(0.00826)	(0.00826)		(0.00827)	(0.00847)	(0.00843)	(0.00825)	(0.00827)	(0.00843)	(0.00843)
Size class: 500+	-0.0389***	-0.0459***	-0.0464***		-0.0434***	-0.0432***	-0.0471***	-0.0437***	-0.0438***	-0.0475***	-0.0468***
	(0.00822)	(0.00827)	(0.00828)		(0.00826)	(0.00851)	(0.00856)	(0.00826)	(0.00825)	(0.00856)	(0.00856)
NUTS1: North-East					-0.00384	-0.00385	-0.00419	-0.00366	-0.00374	-0.00412	-0.00404
					(0.00369)	(0.00371)	(0.00370)	(0.00368)	(0.00368)	(0.00370)	(0.00369)
NUTS1: Centre					0.00454	0.00478	0.00258	0.00430	0.00464	0.00271	0.00269
					(0.00420)	(0.00422)	(0.00421)	(0.00420)	(0.00420)	(0.00421)	(0.00421)
NUTS1: South					0.0293***	0.0296***	0.0270***	0.0274***	0.0290***	0.0267***	0.0260***
					(0.00566)	(0.00569)	(0.00567)	(0.00565)	(0.00565)	(0.00566)	(0.00567)
NUTS1: Islands					0.0269***	0.0272***	0.0249***	0.0235***	0.0263***	0.0244***	0.0229**
					(0.00913)	(0.00915)	(0.00909)	(0.00909)	(0.00910)	(0.00906)	(0.00907)
Liquidity Index						-0.00183**				2.85e-05	7.10e-05
						(0.000906)				(0.000293)	(0.000288)

Table A5 cont. - Perceived credit constraints and firm characteristics: regression results for the period 2016-2019 (a),(b)

VARIABLES	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10	Model 11
Debt Index							0.0808***			0.0781***	0.0705***
							(0.00772)			(0.00768)	(0.00787)
ROA Index								-0.00132***			-0.000875***
								(0.000171)			(0.000176)
ROE Index									-0.000180***	-0.000152***	
									(4.39e-05)	(4.39e-05)	
Observations	21,927	21,927	21,927	21,927	21,927	21,743	21,743	21,927	21,927	21,743	21,743
R-squared	0.006	0.023	0.021	0.020	0.025	0.026	0.032	0.028	0.026	0.032	0.033
Industry Fixed Effects	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region Fixed Effects	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Source: Authors' elaborations on BCS-BR-CCs databases

(a) The table shows results for the linear probability model described in Section 4, where the dependent variable is a dummy equal to 1 if the firm perceives credit constraints each year and 0 otherwise (see Section 2.2), over the period 2016-2019. Missing size class and productivity class coefficients are not reported. Productivity classes are computed within Region-Sector. Regions refer to the 20 Italian regions while sectors refer to the 2-dig NACE Rev. 2 sectoral classification. Reference categories are Year = 2015, Size Class 1-9, and Productivity class 10%. Robust standard errors are in parentheses.

(b) *** p<0.01, ** p<0.05, * p<0.1.

Table A6 - Perceived credit constraints and firm characteristics: regression results for the period 2015-2020 (Logit model) (a), (b)

VARIABLES	Logit Model (LM)	LM - Marginal Eff.
Year: 2016	-0.457***	-0.033***
	(0.0745)	(0.006)
Year: 2017	-0.966***	-0.057***
	(0.0954)	(0.006)
Year: 2018	-0.816***	-0.051***
	(0.0871)	(0.006)
Year: 2019	-0.464***	-0.033***
	(0.0805)	(0.006)
Year: 2020	-0.659***	-0.044***
	(0.0887)	(0.006)
Productivity class: 10%-30%	-0.156	-0.011
	(0.114)	(0.008)
Productivity class: 30%-70%	-0.391***	-0.025***
	(0.110)	(0.008)
Productivity class: 70%-90%	-0.456***	-0.028***
	(0.121)	(0.008)
Productivity class: Top 10%	-0.491***	-0.030***
	(0.119)	(0.008)
Size class: 10-50	-0.334***	-0.021***
	(0.0664)	(0.004)
Size class: 50-250	-0.556***	-0.032***
	(0.0828)	(0.005)
Size class: 250-500	-0.850***	-0.044***
	(0.196)	(0.008)
Size class: 500+	-0.802***	-0.042***
	(0.183)	(0.007)
Debt Index	1.139***	0.063***
	(0.117)	(0.007)
Liquidity Index	-0.00531	-0.000
	(0.00672)	(0.000)
ROA Index	-0.0128***	-0.001***
	(0.00227)	(0.000)
NUTS1: North-East		0.001
		(0.003)
NUTS1: Centre		0.004
		(0.004)
NUTS1: South		0.030***
		(0.005)
NUTS1: Islands		0.026***
		(0.009)
Observations	29,277	29,277
Industry Fixed Effects	Yes	
Pseudo R2	0.0720	

Source: Authors' elaborations on BCS-BR-CCs databases

(a) The table shows results for a Logit model where the dependent variable is a dummy equal to 1 if the firm perceives credit constraints each year and 0 otherwise (Section 2.2). Control variables are those employed in the baseline linear probability model (Section 4 and Table A1). Reference categories are Year = 2015, Size Class 1-9, and Productivity class 10%. Missing size class and productivity class coefficients are not reported. Productivity classes are computed within Region-Sector. Regions refer to the 20 Italian regions while sectors refer to the 2-dig NACE Rev. 2 sectoral classification. Omitted 2-dig FE and macro region FE for LM column and 2-dig FE for Marginal Eff. column. Robust standard errors are in parentheses.

(b) *** p<0.01, ** p<0.05, * p<0.1.

Appendix B - CBS Questionnaire

Q.43. Nowadays, according to you, the conditions for accessing bank credit are more favourable or less favourable compared to those of 3 months ago?
1. More favourable.
2. Constants.
3. Less favourable.
4. Unknown.

Q.44. This judgment stems from your recent direct contacts with banks or financial companies for request/increase the credit of your company or is it just a conviction not linked to specific contacts with banks?
1. It arises from contacts with banks.
2. Independent belief in contacts with banks.

(Only if code 1 to question 44)

Q.45 Have you obtained the credit you requested from the bank or financial company you turned to?
1. Yes, under the same conditions.
2. Yes, but on more expensive terms.
3. No.
4. I had only been to the bank to ask for information.

(Only if code 3 to question 45)

Q.46 The bank or the financial institution did not want to grant /increase the requested credit or you have not accepted the conditions that the bank imposed to grant you the credit (conditions too onerous: interest rates, guarantees, etc.)?
1. The bank did not grant /increase the credit.
2. We have not accepted the conditions that the bank asked for, as it was too expensive.

(Only if code 2 to question 45)

Q.47 What were the main reasons for the aggravation of conditions? (Maximum three answers)

1. Higher rate.
2. More personal guarantees (surety, other contractual obligations).
3. Multiple collateral (physical or financial assets).
4. Limitations for credits granted.
5. Costs (commissions, ancillary costs).

References

Almeida, H., and M. Campello. 2007. “Financial Constraints, Asset Tangibility, and Corporate Investment”. *The Review of Financial Studies*, Volume 20, N. 5: 1429-1460.

Beck, T., A. Demirgüç-Kunt, L. Laeven, and V. Maksimovic. 2006. “The determinants of financing obstacles”. *Journal of International Money and Finance*, Volume 25, N. 6: 932-952.

Berger, A.N., and G.F. Udell. 2006. “A more complete conceptual framework for SME finance”. *Journal of Banking Finance*, Volume 30, N. 11: 2945-2966.

Blanco, R., and N. Jimenez. 2018. “Credit allocation along the business cycle: Evidence from the latest boom bust credit cycle in Spain”. *Working Papers*, N. 1826.

Bottazzi, G., A. Secchi, and F. Tamagni. 2014. “Financial constraints and firm dynamics”. *Small Business Economics*, Volume 42, N. 1: 99-116.

Bottazzi, G., M. Grazzi, A. Secchi, and F. Tamagni. 2011. “Financial and economic determinants of firm default”. *Journal of Evolutionary Economics*, Volume 21, N. 3: 373-406.

Brown, J., S. Fazzari, and B. Petersen. 2009. “Financing Innovation and Growth: Cash Flow, External Equity, and the 1990s RandD Boom”. *Journal of Finance*, Volume 64, N. 1: 151-185.

Canton, E., I. Grilo, J. Monteagudo, and P. van der Zwan. 2013. “Perceived credit constraints in the European Union”. *Small Business Economics*, Volume 41, N. 3: 701-715.

Cosh, A., D. Cumming, and A. Hughes. 2009. “Outside Entrepreneurial Capital”. *The Economic Journal*, Volume 119, N. 540: 1494-1533.

Devereux, M., and F. Schiantarelli. 1990. “Investment, financial factors and cash flow from U.K. panel data”. In Hubbard, G. *Information, capital markets and investment*: 279-306. Chicago, IL, U.S.: University of Chicago Press.

Djankov, S., C. McLiesh, and A. Shleifer. 2007. "Private credit in 129 countries". *Journal of Financial Economics*, Volume 84, N. 2: 299-329.

Farre-Mensa, J., and A. Ljungqvist. 2016. "Do Measures of Financial Constraints Measure Financial Constraints?". *The Review of Financial Studies*, Volume 29, N. 2: 271-308.

Fazzari, S., R.G. Hubbard, B.C. Petersen, A.S. Blinder, and J.M. Poterba. 1988. "Financing Constraints and Corporate Investment". *Brookings Papers on Economic Activity*, Volume 19, N. 1: 141-206.

Ferrando, A., and K. Mulier. 2022. "The real effects of credit constraints: Evidence from discouraged borrowers". *Journal of Corporate Finance*, N. 73.

Ferrando, A., and N. Griesshaber. 2011. "Financing Obstacles Among Euro Area Firms: Who Suffers the Most?". *ECB Working Paper*, N. 1293.

Finnegan, M., S. and Kapoor. 2023. "ECB unconventional monetary policy and SME access to finance". *Small Business Economics*, Volume 61, N. 3: 1253-1288.

Gertler, M. 1988. "Financial structure and aggregate economic activity: An overview". *Journal of Money, Credit, and Banking*, Volume 20, N. 3: 559-596.

Gómez, M.G.-P. 2019. "Credit constraints, firm investment and employment: Evidence from survey data". *Journal of Banking and Finance*, Volume 99: 121-141.

Hadlock, C.J., and J.R. Pierce. 2010. "New Evidence on Measuring Financial Constraints: Moving Beyond the KZ Index". *The Review of Financial Studies*, Volume 23, N. 5: 1909-1940.

Hall, B.H. 2002. "The Financing of Research and Development". *Oxford Review of Economic Policy*, Volume 18, N. 1: 35-51.

Holton, S., M. Lawless, and F. McCann. 2013. "SME financing conditions in Europe: Credit crunch or fundamentals?". *National Institute Economic Review*, Volume 225, N. 1: 52-67.

Istituto Nazionale di Statistica - Istat. 2023. *Rapporto sulla competitività dei settori produttivi. Edizione 2023*. Roma, Italia: Istat.

Istituto Nazionale di Statistica - Istat. 2021. *Rapporto sulla competitività dei settori produttivi. Edizione 2021*. Roma, Italia: Istat.

Jaffee, D., and J.E. Stiglitz. 1990. "Credit rationing". In Friedman, B.M., and F.H. Hahn (*eds*.). *Handbook of Monetary Economics. Volume 2,* Chapter 16: 837-888. Amsterdam, The Netherlands: Elsevier.

Jimenez, G., S. Ongena, J.-L. Peydro, and J. Saurina. 2012. "Credit supply and monetary policy: Identifying the bank balance-sheet channel with loan applications". *American Economic Review*, Volume 102, N. 5: 2301-2326.

Kaplan, S.N., and L. Zingales. 2000. "Investment-Cash Flow Sensitivities Are Not Valid Measures of Financing Constraints". *The Quarterly Journal of Economics*, Volume 115, N. 2: 707-712.

Kaplan, S.N., and L. Zingales. 1997. "Do Investment-Cash Flow Sensitivities Provide Useful Measures of Financing Constraints". *The Quarterly Journal of Economics*, Volume 112, N. 1: 169-215.

Kuntchev, V., R. Ramalho, J. Rodríguez-Meza, and J.S. Yang. 2016. "What Have We Learned from the Enterprise Surveys Regarding Access to Credit by SMEs?". *World Bank Policy Research Working Paper*, N. 6670.

Lahr, H., and A. Mina. 2020. "Endogenous financial constraints and innovation". *Industrial and Corporate Change*, Volume 30, N. 3: 587-621.

McQuinn, J. 2019. "SME access to finance in Europe: Structural change and the legacy of the crisis". *Central Bank of Ireland Research Technical Paper*, Volume 2019, N. 10.

Michelacci, C., and V. Quadrini. 2009. "Financial Markets and Wages". *The Review of Economic Studies*, Volume 76, N. 2: 795-827.

Minetti, R., and S. Chun Zhu. 2011. "Credit constraints and firm export: Microeconomic evidence from Italy". *Journal of International Economics*, Volume 83, N. 2: 109-125.

Musso, P., and S. Schiavo. 2008. "The impact of financial constraints on firm survival and growth". *Journal of Evolutionary Economics*, Volume 18, N. 2.

Revest, V., and A. Sapio. 2012. "Financing technology-based small firms in Europe: what do we know?". *Small Business Economics*, Volume 39, N. 1: 179-205.

Secchi, A., F. Tamagni, and C. Tomasi. 2016. "Export price adjustments under financial constraints". *The Canadian Journal of Economics*, Volume 49, N. 3: 1057-1085.

Stiglitz, J.E., and A. Weiss. 1981. "Credit Rationing in Markets with Imperfect Information". *The American Economic Review*, Volume 71, N. 3: 393-410.

Storey, D.J. 2016. *Understanding The Small Business Sector*. London, UK: Routledge.

Whited, T.M., and G. Wu. 2006. "Financial Constraints Risk". *The Review of Financial Studies*, Volume 19, N. 2: 531-559.

The integrated system of statistical-economic classifications: an extension to national versions of the NACE

Francesca Alonzi[1]

Abstract

An integrated system of international statistical economic classifications gradually emerged during the 1980s. Primarily developed under the guidance of the United Nations Statistical Division, this system includes interconnected classifications for activities, products, and goods. It includes international or reference classifications, as well as regional standards (e.g. European) and derived national versions. The development and maintenance of pairwise links between two national classifications derived from the same European classification (e.g. the NACE Rev. 2) is not common practice. However, it would have some advantages for European stakeholders. This research aims to investigate the direct connections between the Italian ATECO classification and the Swiss NOGA classification of economic activities. The goal is to facilitate the development and updating of correspondence between these classifications, as well as with the national versions of NACE.

Keywords: Economic activities, NACE, Official Statistics, Statistical classifications.

JEL Classification: C18, F00.

DOI: https://doi.org/10.1481/ISTATRIVISTASTATISTICAUFFICIALE_1-2-3.2023.04.

1 Francesca Alonzi (alonzi@istat.it), Italian National Institute of Statistics - Istat.

The views and opinions expressed are those of the author and do not necessarily reflect the official policy or position of the Italian National Institute of Statistics - Istat.

The author would like to thank the anonymous reviewers for their comments and suggestions, which enhanced the quality of this article.

1. Introduction[2]

An integrated system of international statistical economic classifications, also known as "international system of economic classifications" or as "integrated system of statistical activity and product classifications", gradually emerged during the 1980s (UNSD, 2008). Developed under the auspices of the United Nations Statistical Division, this system comprises activity, product, and goods statistical classifications that are interconnected to one another. It encompasses international, regional and national classifications; the latter is based on the former but has been adapted to suit national economies.

National classifications have to be consistent with their reference classifications at the global and regional level to ensure coherence and comparability between statistics on a global scale (the so-called comparability across countries). Sometimes such a consistency is regulated by legal frameworks, as in the case of national versions of the European statistical classification of economic activities NACE Rev. 2 (Eurostat, 2008), whose conformity to the NACE has to be checked by the Commission.

The coding system and the structure (coding system and titles/descriptions of each category) represent the core elements of a statistical classification and the simplest way to link different classifications; when two classifications share part of their structures (*e.g.* the highest levels), statistical data produced according to them meet a certain degree of consistency. Nevertheless, even when two classifications are not linked via the structure, they may be linked by correspondence tables developed *ad hoc* by experts in this field.

All statistical classifications belonging to the integrated system of statistical activity and product classifications are linked via the structure or correspondence tables. Over time, the connections between statistical classifications have become increasingly accurate due to the skills and expertise gained by statisticians. Additionally, new information systems have been developed to help organise metadata and various types of information, which are essential for correctly interpreting statistical classifications.

In this context, the development (and maintenance) of pairwise links between two national classifications derived from the same European

2 The author would like to thank Marc Froidevaux and Oliver Gallusser (Swiss Federal Statistical Office) for their comments and suggestions, which enhanced the quality of this article.

classification (*e.g.* the NACE Rev. 2) is not common practice, especially because of the high costs that are necessary to work on it. Nevertheless, such a practice would have several advantages for European stakeholders and a significant impact on the processes of maintenance, update and revision of national and European classifications of economic activities. For instance, it would help check the European relevance of an economic activity (the number of European countries that have developed the same categories at the national level, differently from the NACE). At the same time, it could be used to investigate and verify the degree of consistency between national classifications directly derived from the NACE. The informative potential derived from the development of pairwise links between two national versions of the same statistical classification may also encourage drafting best practices and guidelines to support other Member States. Finally, European and national stakeholders would benefit from the availability of comparable national data, referred to in some countries, which would result in being more detailed than that available at the European level.

This research work aims to explore direct links between the Italian ATECO and the Swiss NOGA (General Classification of Economic Activities) to compare the two schemes and highlight best practices with different solutions to derive national versions of the NACE. In this respect, an extension to national versions of the NACE is provided to the integrated system of statistical economic classifications.

The contents of this work are divided into six Sections, including the first one aimed at introducing the work. The second Section is devoted to introducing the issues dealt with in this work by describing some basic principles of statistical classifications and the reasons behind the existence of an integrated system of statistical economic classifications. The third Section is focussed on the data and methods used to reach the research objectives, that is, examining the differences between the ATECO and the NOGA and thus identifying possible interlinkages between national versions of the NACE. The fourth Section presents the main results by providing comparison case studies between the two classifications. Finally, the fifth Section provides some conclusive remarks as well as some possible future developments of the work done so far. The complete correspondence table between the two national versions of the NACE classification is provided in the Appendix.

2. Background

According to our knowledge and experience in this field, the literature on statistical classifications is still scarce; such a circumstance is due especially to two main factors.

First of all, standard statistical classifications, *e.g.* those of economic activities, are relatively young. The NACE classification, for example, has its roots in the "*Nomenclature des Industries établies dans les Communautés Européennes*" (NICE) (Classification of Industries Established in the European Communities) that was developed in the 1960s (Eurostat, 2008).

Secondly, statistical classifications represent a niche topic that needs expert statisticians having several cross-skills to be maintained, updated and revised. Statistical classifications were born to organise and present statistics (Hoffman and Chamie, 1999), so they are based on conceptual and methodological frameworks as well as on practical considerations concerning data collection and compilation. Such a feature makes classifications play a central role in the statistical system as a whole. If the wrong categories are applied to the observed universe, the information supplied to decision-makers for their analysis will be of poor quality or even useless.

2.1 The integrated system of economic classifications

An integrated system of international statistical economic classifications, also known as "international system of economic classifications" or as "integrated system of statistical activity and product classifications", gradually emerged during the 1980s.

Graphically speaking, this integrated system can be presented as a matrix containing various classifications, which are conceptually related, organised within the matrix itself according to the two key dimensions listed below.

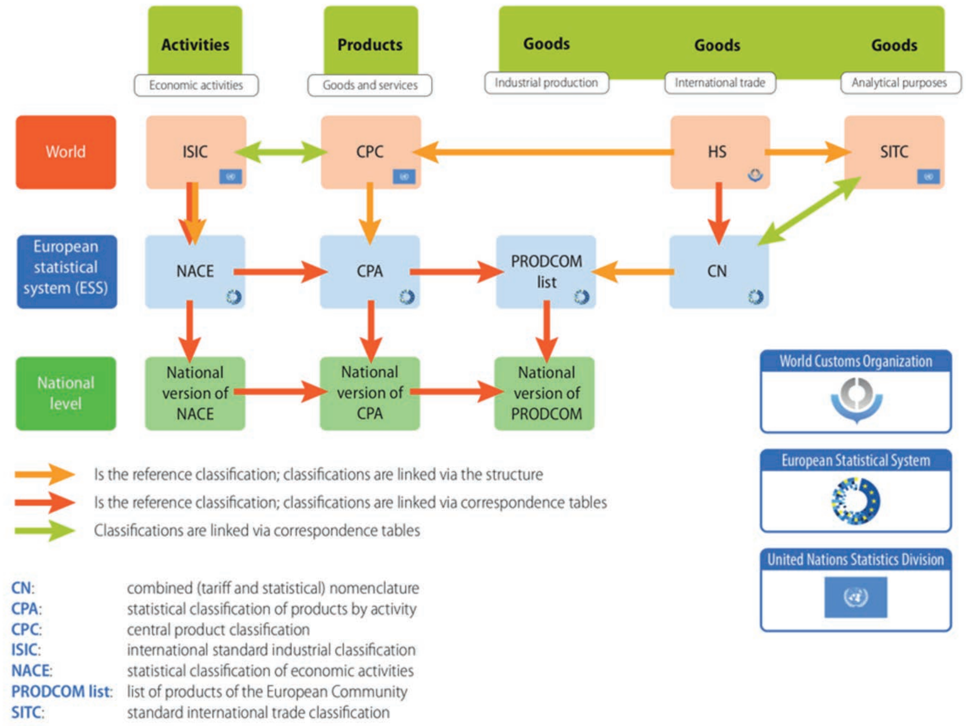
1. The family of each classification included in the system: economic activities, products (goods and services) and goods (organised as industrial production, international trade and analytical purposes). Economic activity classifications are listed in the first column, followed by product classifications, and finally goods classifications on the right side of the scheme.

2. The geographical scope of each classification included in the system: world, regional and national level. Classifications used at the world level are placed in the first row at the top of the matrix, while regional standards (*e.g.* European) occupy the middle row, and national versions are presented in the bottom row.

The system is essentially a set of classifications connected by horizontal and vertical relationships. Horizontal relationships ensure comparability between classifications belonging to different families (economic activities, products and goods). On the other hand, vertical relationships identify classifications belonging to the same family but referring to different geographical scopes: world, regional, and national levels.

Figure 2.1 illustrates the whole system adapted to the European level.

Figure 2.1 - The integrated system of statistical economic classifications adapted to the European level



Source: Eurostat, 2021

In Figure 2.1, national versions of the European NACE classification are thus placed in the bottom corner on the left of the matrix. Being at the lowest levels, national classifications inherit the methodological framework and rules established at the higher levels.

Belonging to such a system, data released according to national versions are comparable with those produced internationally according to the other classifications of the system, thanks to the existence of correspondence relationships and hierarchical constraints (related to the structure).

Links via correspondence tables or via the structure within the entire system have an impact on the room for change within an element of the system (they can be seen as "structural constraints") and on the timeline that has to be followed ("process constraints").

Structural constraints mean that changes applied to a classification in the system are only those that satisfy the essential relationships with its reference classification; for example, the NACE classification is a derived classification of the ISIC (International Standard Industrial Classification): categories at all levels of the NACE are defined either to be identical to, or to form subsets of, single ISIC categories.

Process constraints concern the timing of revision processes of classifications within the system: if the NACE and the ISIC are linked via the structure, changes in the structure of the ISIC (as the reference classification) should be applied before or during the revision process of the NACE.

Within the system, structural and process constraints exist because "*national and international classifications are mutually dependent*" (Hoffman and Chamie, 1999).

In 1999, Hoffmann and Chamie wrote: "*Linking to a common reference classification when only two countries are involved, the need for comparable information can be satisfied most effectively by directly linking their national classifications […]*".

By exploiting the relationships within the integrated system of statistical activity and product classifications, this research aims to explore direct links between the Italian ATECO (Istat, 2022) and the Swiss NOGA classifications of economic activities to compare the two schemes and identify best practices, as well as different solutions to derive national versions of the NACE.

The potential of the integrated system of economic classifications and the application of correspondence tables between different classifications have also been recently presented in relation to the the COVID-19 emergency (Alonzi *et al.*, 2020 *a* and *b*).

2.2 Correspondence tables: an introduction

Though different, classifications within the system are closely linked. Linkage is achieved either through the coding system used in the classifications or by means of correspondence tables (Eurostat, 2021).

A correspondence table (also known as a concordance table, mapping, or correlation table) consists of establishing links between the codes in a source classification with the corresponding codes in a target classification (Eurostat, 2019), detailing how a category in one classification relates or links to the new/other classification (Hancock, 2013).

There exist two types of correspondence tables:

- providing links between two versions of the same statistical classification, such as NACE Rev. 1.1 and NACE Rev. 2;
- providing links between two distinct statistical classifications belonging to the same family, such as the CPC (Central Product Classification) and the CPA (Statistical Classification of Products by Activity), which are both product classifications or belonging to different families, such as the ISIC and the CPC.

As perfectly described by Eurostat (2019), a correspondence table can consist of the following relationships:

- one-to-one (1:1) → the whole content of a position in the source classification corresponds exactly to the whole content of a position in the target classification (even if the textual labels are different);
- one-to-many (1:n) → the content of a position in the source classification is distributed over more than one position in the target classification;
- many-to-one (m:1) → the content of several positions in the source classification is grouped into a single position in the target classification;
- many-to-many (m:n) → *m* number of positions in the source classification corresponds to *n* number of positions in the target classification.

It is not an easy task to organise each link in one of the above types; it could be useful to set methodological constraints, deciding if all kinds of links can be defined or not. Depending on the classifications involved in the correspondence, such lists of links can also have attributes; the role of attributes is to support users in better understanding the relationship between the two classifications involved. Eurostat (2019) has recently mapped the possible attributes: specification of partial coverage, comment and information on the status of the links.

For what we are interested in, the specification of partial coverage is extremely important to present correspondences to users. This attribute, which can be used for both the source and target classifications, is used to indicate that only a part of the position concerned is included in a specific link and that the rest of the content of the position is to be found in one or more other links.

The definition of correspondences is mandatory between international classifications, as they facilitate international reporting and enable time-series management (Hancock, 2013). However, the mapping of different classifications to each other is also dependent on user needs.

Sometimes, possible links between classifications may be very complicated. The correspondence tables should be presented in a form that is technically accurate and easy to understand for users. This will typically require separate tables for each direction (Hoffmann and Chamie, 1999).

In this research work, correspondences are presented in the Appendix. Each pairwise link is presented in a row and the following information are provided: classification A (the code of the source classification), coverage A (the type of coverage - total or partial - referred to code A), classification B (the code of the target classification), coverage B (the type of coverage - total or partial - referred to code B) and the type of link (one-to-one, one-to-many, many-to-one, many-to-many).

3. Data and methods

3.1 Data

The Italian and Swiss classifications of economic activities (structure and contents), namely ATECO and NOGA, are the core elements of this research work. They represent the national versions of the European NACE classification, reflecting national circumstances and priorities than the European ones; at the moment of writing, NACE Rev. 2 is the current version.

From a structural point of view, the NACE includes four levels (sections, divisions, groups and classes). Legislation at the European Union level[3] establishes the uniform use of the classification across all Member States. Based on the NACE Regulation, Member States' statistics presented according to economic activities are to be produced using NACE Rev. 2 or a national classification derived therefrom. The national classification may introduce additional headings and levels, and a different coding may be used. Each of the levels, except for the highest, shall consist of either the same headings as the corresponding NACE Rev. 2 level or headings constituting an exact breakdown thereof.

Several countries, such as Austria, France and Italy, have national versions of the NACE; Switzerland has also adopted a national version. The reason lies in the evidence that the NACE does not entirely satisfy national needs and requires an adaptation to meet countries' specific conditions. Sometimes, national versions of the NACE, developed primarily for the production and presentation of statistics, are also utilised for other purposes, *e.g.* legal purposes; thus, further breakdowns are necessary to identify economic activities in a more delimited way.

In general, but not always, national versions of the NACE are more detailed than their European version and include at least one more level.

The Italian version of the NACE includes two additional levels (categories and sub-categories) and attributes several national positions, which are greater in number than the codes found in the Swiss NOGA. The key differences between the classifications are summarised in Table 3.1 below. These

3 The legal framework is provided by Regulation (EC) N. 1893/2006 of the European Parliament and of the Council of 20 December 2006 establishing the statistical classification of economic activities NACE Revision 2 and amending Council Regulation (EEC) N. 3037/90 as well as certain EC Regulations on specific statistical domains.

differences relate to the number and names of the classification levels, their positions within each level, the additional levels at the national level, and the coding systems used by each country.

Table 3.1 - A comparison between two national versions of the NACE: ATECO and NOGA

<table>
<tr><th>Name</th><th>Current version</th><th>Country</th><th>Level</th><th>Name of extra levels</th><th>No. of positions</th><th>Coding system</th></tr>
<tr><td rowspan="2">ATECO</td><td rowspan="2">ATECO 2007
2022 update</td><td rowspan="2">Italy</td><td>5</td><td>Categories</td><td>920</td><td>XX.XX.Y
→ Y being a number between 0 and 9</td></tr>
<tr><td>6</td><td>Sub-categories</td><td>1,241</td><td>XX.XX.YZ
→ Z being a number between 0 and 9</td></tr>
<tr><td>NOGA</td><td>NOGA 2008</td><td>Switzerland</td><td>5</td><td>Types</td><td>794</td><td>XXXXYY
→ YY being a number between 00 and 99</td></tr>
</table>

Source: Author’s processing of Istat, 2022, and Swiss Federal Statistical Office, 2008

In this research, sub-categories of *ATECO 2007 aggiornamento* (update) *2022* (hereafter referred to as *ATECO 2022*) and types of *NOGA 2008* have been compared (Istat, 2009 and 2022; Swiss Federal Statistical Office, 2008).

The ATECO classification was developed by the Italian National Institute of Statistics (Istat) in collaboration with experts from administrative bodies involved in the registration, including the classification, of economic units, as well as representatives of numerous trade associations. Even if the current ATECO came into force on January 1, 2008, it was finalised in 2007 and thus called “ATECO 2007”. It was realised in parallel with the NACE Rev. 2; as a national version of the NACE, it is fully consistent with the four-digit structure defined at the European level, but it has two more digits. More specifically, while the first four levels are inherited from the NACE, two additional levels have been introduced at the national level: “categories” and “sub-categories”. In most cases, ATECO categories and sub-categories consist of a conversion of the notes of inclusion described in the NACE into national headings (Alonzi and Viviano, 2022).

The ATECO is a hierarchical classification; its structure coding pattern can be summarised as XX.XX.YZ where XX.XX denotes NACE codes, Y denotes extra codes added for statistical purposes at the national level (categories), and Z denotes extra codes added for other than statistical aims (sub-categories).

Since the release of the NACE Rev. 2, the corresponding ATECO classification in force (ATECO 2007) has been used at the national level both for statistical and non-statistical purposes. It is worth noting that the sixth level is not used for statistical purposes, but it is limited to administrative functions. More specifically, in the Statistical Business Registers maintained by Istat, economic units are classified based on the ATECO categories rather than the sub-categories. On the contrary, in administrative sources (such as fiscal sources or the register of enterprises managed by the Chamber of Commerce), enterprises are registered at the highest level of detail according to the ATECO sub-categories. ATECO is also widely used by the government local agencies, trade associations, and other organisations. The ATECO 2007 classification was updated by Istat with the support of the ATECO Committee[4] in 2021 (*ATECO 2007 aggiornamento 2021*) and in 2022 (*ATECO 2007 aggiornamento 2022*).

The complete ATECO 2007 *aggiornamento* 2022 (structure and explanatory notes) is available only in Italian at the Istat's website (www.istat.it).

The General Classification of Economic Activities (NOGA) is derived from the NACE. The NOGA was developed by the Swiss Federal Statistical Office (FSO) in collaboration with experts from the public administration and numerous umbrella organisations, taking into account the needs of various interest groups in Switzerland.

The particularities and/or important activities of the national economy have been incorporated in the last two positions of the six-digit code, which represents the fifth level of the classification called "type". As a consequence, the current version of the NOGA has a five-level hierarchical structure: section, division, group, class (4 digits) and type (6 digits). The NOGA is compatible with the NACE up to the fourth level (class).

The NOGA coding pattern can be summarised as XXXXYY, where XXXX denotes NACE codes without the "." separating the second and the third digit, while YY denotes extra codes added for the NOGA. The last level of the NOGA has 794 positions. When the national breakdown is not relevant, the last digits are "00".

4 The ATECO Committee was first set up in June 2020 as a temporary organism, although it is a candidate to become permanent in the future; it is coordinated by Istat and formed by representatives of statistical domains, administrative bodies and federations/associations of industries. The primary goal of the ATECO Committee is to establish a revised version of the ATECO classification by consulting with producers and users of statistics, as well as experts in relevant subject areas.

The current NOGA 2008 came into force on 1st January 2008 and has been integrated into the Business and Enterprise Register (BER); a NOGA code is assigned to each enterprise and each business registered in the BER. From a statistical point of view, the NOGA makes it possible to classify companies according to their main activity and to group them into coherent units. It is thus used as a basis for many economic statistics. From an administrative point of view, insurance companies, employment agencies, pension funds, etc., use NOGA codes on their initiative (*e.g.* to determine their risk premiums). Nevertheless, the NOGA was designed for statistical purposes and is not directly linked to any legal basis. Therefore, the use of the NOGA in a non-statistical context is not the responsibility of the FSO.

The complete NOGA 2008 (structure and explanatory notes) is available in four languages (English, French, Italian and German) at www.bfs.admin.ch.

3.1 Methods

To compare two national versions of the same European classification of economic activities, a complete correspondence table between the ATECO and the NOGA has been developed.

All the steps applied to derive pairwise links at the highest level of detail (ATECO sub-categories ↔ NOGA types) are summarised in Table 3.2 and described below.

Table 3.2 - Overview of the methodological steps followed to define pairwise links between the ATECO and the NOGA classifications

Step	Description
1	Defining the structure (bidirectional links), constraints (no constraints set) and attributes (total or partial coverage) of the links.
2	Detection of automatic one-to-one (1:1) links meeting the condition "one NACE code = one ATECO code = one NOGA code". Only the structure of the ATECO and NOGA classifications has been taken into account.
3	Detection of automatic one-to-many (1:n) links meeting the condition "one NACE code = one ATECO code = more than one NOGA code". Only the structure of the ATECO and NOGA classifications has been taken into account.
4	Detection of automatic many-to-one (m:1) links meeting the condition "one NACE code = more than one ATECO code = one NOGA code". Only the structure of the ATECO and NOGA classifications has been taken into account.
5	Within the remaining codes, detection of manual one-to-one (1:1), one-to-many (1:n) and many-to-one (m:1) links other than automatic links already identified in steps 2, 3 and 4. Explanatory notes of the ATECO and NOGA classifications have been consulted.
6	Flagging the remaining links as many-to-many (m:n) links and setting the pairwise correspondences. Explanatory notes of the ATECO and NOGA classifications have been consulted.
7	Applying *ad hoc* checks to verify the derived links.
8	Results analysis and systematisation.

Source: Author's processing

Step 1. In the first step, it was decided how to structure and present the links. The derived correspondence table is bidirectional; thus, it can be read in two ways: "ATECO sub-categories to NOGA types" and "NOGA types to ATECO sub-categories", thanks to the specification of partial coverage both from the ATECO and the NOGA perspectives. This attribute is used to indicate that only a part of the position concerned is included in a specific link and that the rest of the content of the position is to be found in one or more other links.

No constraints have been set; despite that, ideally, links could have been possible (thus, valid) only within the same NACE class; avoiding such a constraint made it possible to discover some inconsistencies between the ATECO and the NOGA concerning the NACE.

Step 2. Automatic one-to-one (1:1) links meeting the following condition have been created: a NACE code corresponds to only one ATECO code and only one NOGA code. Such a condition is valid when the ATECO code ends with "00" and the last two digits of the NOGA code are "00". To derive such a type of link, only the detailed structures of the two classifications have been taken into account; explanatory notes have not been investigated.

Step 3. Automatic one-to-many (1:n) links meeting the following condition have been set: a NACE code corresponds to only one ATECO code, but to many NOGA codes. This means that the final digits of the ATECO sub-category are always equal to "00", but the NOGA type does not end with "00". To derive such a type of link, only the detailed structures of the two classifications have been taken into account.

Step 4. Automatic many-to-one (m:1) links meeting the following condition have been identified: a NACE code corresponds to many ATECO sub-categories and only to one NOGA code. This means that the final digits of the ATECO sub-category are not equal to "00" while the NOGA type ends with "00". To derive such a type of link, only the detailed structures of the two classifications have been taken into account.

Step 5. Other one-to-one (1:1), one-to-many (1:n), and many-to-one (m:1) links have been set within the remaining codes not already included in the above-mentioned conditions (steps 2, 3 and 4). A thorough examination of the explanatory notes for the ATECO and NOGA classifications was necessary to complete this step.

Step 6. The remaining links have been flagged as many-to-many (m:n) links, and the pairwise correspondences have been created, making use of explanatory notes of the ATECO and NOGA classifications. Many-to-many links are those that meet the following condition: a NACE code corresponds to many ATECO sub-categories and many NOGA codes, and many ATECO sub-categories correspond to many NOGA codes.

Step 7. Two types of ad hoc checks have been applied to verify the derived links (the so-called candidate correspondence table): manual conceptual checks to verify if the clusters contain conceptually related content, and automatic checks based on the type of coverage (total or partial); the latter are presented in Table 3.3.

Step 8. Once all the possible links have been identified, they have been analysed and the results systematised.

Table 3.3 - Step 7: checking the derived links between the ATECO (A) and the NOGA (B) classifications

Condition	Statement (type of linkage)
coverage A = total and coverage B = total	one-to-one
coverage A = partial and coverage B = total and no other codes with coverage B = partial in the same cluster	one-to-many
coverage A = total and coverage B = partial and no other codes with coverage A = partial in the same cluster	many-to-one
coverage A = partial and coverage B = partial or coverage A = partial and coverage B = total and other codes with coverage B = partial in the same cluster or coverage A = total and coverage B = partial and other codes with coverage A = partial in the same cluster	many-to-many

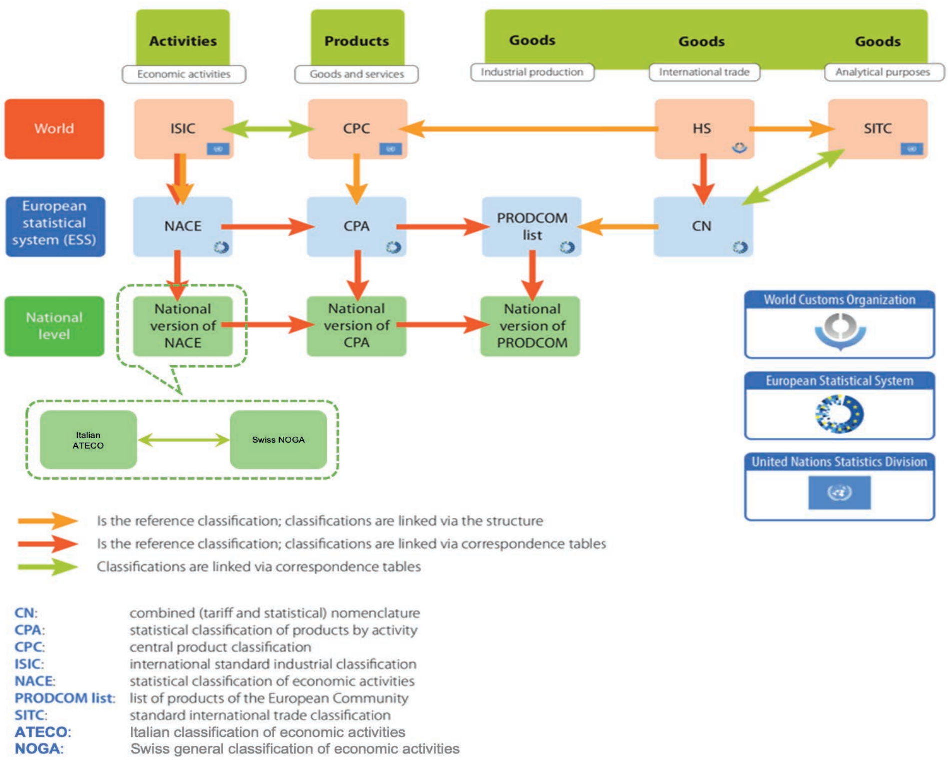
Source: Author’s processing

4. Main results

This Section presents the main results of the research work by providing comparison case studies between the ATECO and the NOGA classifications. Consistency errors as well as difficulties in the interpretation of the same statistical classification are shown at the end of this Section, too. The complete correspondence table is provided as an Appendix.

In this respect, an extension to national versions of the NACE is provided to the integrated system of statistical economic classifications. It is presented as a focus on the lowest level of the system, which is dedicated to national versions, and it shows the ATECO and the NOGA classifications separately but linked via a correspondence table. The reference classification of economic activities for both of them is always the NACE, and moving up, the ISIC.

Figure 4.1 - The integrated system of statistical economic classifications extended to ATECO and NOGA



Source: Author’s processing of Eurostat, 2021

A total of 1,397 pairwise links between the ATECO and the NOGA have been identified, as presented in Table 4.1. Many-to-one links (a NOGA type corresponds to at least two ATECO sub-categories) represent around 50% of the total number of links, while one-to-one links are less than 30% of the total number of links.

Table 4.1 – Number of links between ATECO 2022 and NOGA 2008 by type

TYPE OF LINK	Description	No. of links	%
one-to-one	An ATECO sub-category corresponds to only one NOGA type	396	28.35
one-to-many	An ATECO sub-category corresponds to at least two NOGA types	118	8.45
many-to-one	A NOGA type corresponds to at least two ATECO sub-categories	698	49.96
many-to-many	Many ATECO sub-categories correspond to many NOGA types	185	13.24
TOTAL		**1,397**	**100.00**

Source: Author's processing

4.1 One-to-one (1:1) links

The links of one-to-one (1:1) type are 396; they may be divided into at least two main categories: automatic links and manual links (Table 4.2).

Table 4.2 – Links of one-to-one (1:1) type by method (automatic or manual)

TYPE OF LINK	No. of links	%
Automatic one-to-one links: No national breakdown is provided for Italy or Switzerland; the European NACE classification is used without further breakdowns.	343	86.62
Manual one-to-one links: Some breakdown is provided at the national level.	53	13.38
TOTAL	**396**	**100.00**

Source: Author's processing of Istat, 2022, and Swiss Federal Statistical Office, 2008

In the first case, no national breakdown is provided at the Italian or Swiss level: the European NACE classification is used at both the Italian and Swiss levels without any further breakdown; 343 links meet this condition. It should be noted that in some cases, slight differences in the titles have been observed.

The following examples illustrate one-to-one links.

Activities of growing of oleaginous fruits:

ATECO 01.26.00 *Coltivazione di frutti oleosi* ↔ NOGA 012600 *Coltivazione di frutti oleosi*

Activities of raising swine/pigs:

ATECO 01.46.00 *Allevamento di suini* ↔ NOGA 014600 *Allevamento di suini*

Activities of extraction of natural gas:

ATECO 06.20.00 *Estrazione di gas naturale* ↔ NOGA 062000 *Estrazione di gas naturale*

Activities of reinsurance:

ATECO 65.20.00 *Attività di riassicurazione* ↔ NOGA 652000 *Riassicurazioni*

Activities of fitness facilities:

ATECO 93.13.00 *Gestione di palestre* ↔ NOGA 931300 *Istituti di ginnastica e fitness*

On the contrary, some breakdown is provided at the national level in 53 links. Some examples are provided below.

Activities of embroidery:

ATECO 13.99.10 *Fabbricazione di ricami* ↔ NOGA 139901 *Fabbricazione di ricami*

Activities of surveyors:

ATECO 71.12.30 *Attività tecniche svolte da geometri* ↔ NOGA 711204 *Studi di geometri*

Activities of language schools:

ATECO 85.59.30 *Scuole e corsi di lingua* ↔ NOGA 855901 *Istruzione linguistica*

Activities of physiotherapy:

ATECO 86.90.21 *Fisioterapia* ↔ NOGA 869002 *Fisioterapia*

Activities of freelance journalists:

ATECO 90.03.01 *Attività dei giornalisti indipendenti* ↔ NOGA 900303 *Giornalisti indipendenti*

In all one-to-one (1:1) links, Italian data produced according to the ATECO classification can be easily compared with Swiss data produced according to the NOGA classification.

4.2 One-to-many (1:n) links

The correspondence table also includes 118 **links of one-to-many (1:n) type**, meaning that an ATECO sub-category corresponds to at least two NOGA types.

In 74 1:n links, an ATECO sub-category corresponds to at least 3 NOGA types (Table 4.3).

There is even a 1:11 correspondence to describe some activities classified within NACE Rev. 2 class 64.19 "Other monetary intermediation" and ATECO sub-category 64.19.10, where the NOGA provides the following types:

641901 Banks with a special field of business

641902 Cantonal banks

641903 Big banks

641904 Regional banks and savings banks

641905 Raiffeisen banks

641906 Commercial banks

641907 Stock Exchange banks

641908 Foreign-controlled banks

641909 Branches of foreign banks

641910 Private bankers

641911 Other banking institutions

Table 4.3 – One-to-many links between ATECO 2022 and NOGA 2008

No. of ATECO sub-categories	No. of NOGA types	No. of clusters	No. of links
1	2	22	44
1	3	14	42
1	4	4	16
1	5	1	5
1	11	1	11
		42	**118**

Source: Author's processing

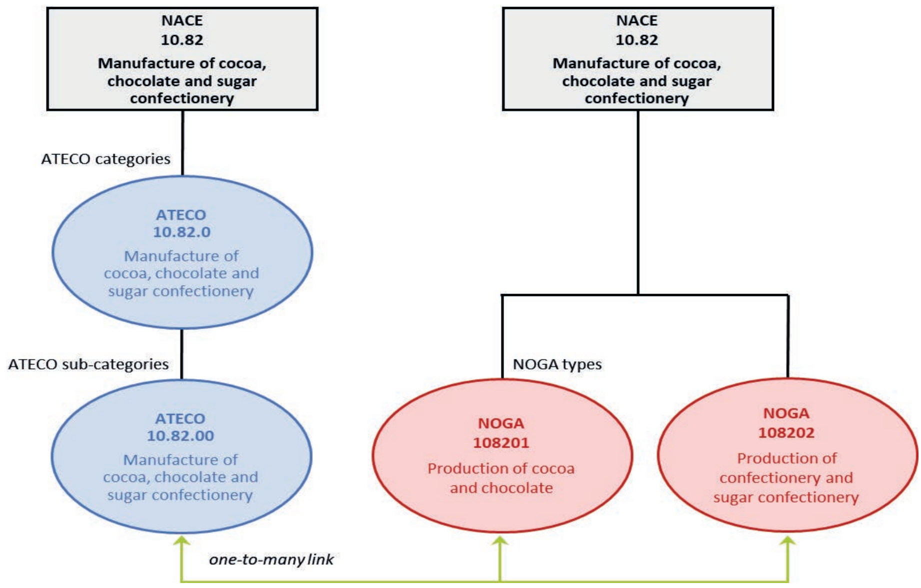
Among one-to-many links, an example concerns NACE Rev. 2 class 10.82 "Manufacture of cocoa, chocolate and sugar confectionery" that does not have a national breakdown at the Italian level (ATECO category 10.82.00

Produzione di cacao in polvere, cioccolato, caramelle e confetterie) while it has been split into two NOGA types in Switzerland (NOGA 108201 "Production of cocoa and chocolate" and NOGA 108202 "Production of confectionery and sugar confectionery"). As a consequence, the correspondence table includes two links: the first one between ATECO 10.82.00 and NOGA 108201 and the other between ATECO 10.82.00 and NOGA 108202.

ATECO 10.82.00 ↔ NOGA 108201

ATECO 10.82.00 ↔ NOGA 108202

Figure 4.2 - ATECO and NOGA national links within NACE class 10.82



Source: Author's processing of Istat, 2022, and Swiss Federal Statistical Office, 2008

Another example of this type of link is the retail sale of glasses and other optical goods, as well as photographic articles. At the Italian level these activities are classified together in ATECO code 47.78.20 *Commercio al dettaglio di materiale per ottica e fotografia*, while they are split into two different codes in the NOGA: 477802 *Commercio al dettaglio di occhiali e articoli simili* and 477803 *Commercio al dettaglio di articoli fotografici*. As a consequence, the correspondence table includes two links: the first one between ATECO 47.78.20 and NOGA 477802 and the other between ATECO 47.78.20 and NOGA 477803.

ATECO 47.78.20 ↔ NOGA 477802

ATECO 47.78.20 ↔ NOGA 477803

Being more precise, the specification of partial coverage on the ATECO side could be presented as well to support the user in understanding the kind of relationship existing between the two classifications:

Partial ATECO 47.78.20 ↔ *Total* NOGA 477802

Partial ATECO 47.78.20 ↔ *Total* NOGA 477803

The two clusters introduced above differ in that ATECO 47.78.20 does not correspond exactly to NACE class 47.78, while ATECO 10.82.00 corresponds exactly to NACE class 10.82. In the first case, Italy reaches a national breakdown at the level of category (5-digit-level) instead of that of sub-category; this is evident from the fact that the ATECO code contains "2" instead of "0" in the fifth position 47.78.20. In effect, ATECO 47.78.20 is derived from the national-level split of NACE class 47.78 into seven categories (47.78.1, 47.78.2, 47.78.3, 47.78.4, 47.78.5, 47.78.6, and 47.78.9), corresponding to seven economic activities. This choice is because the 5-digit-level ATECO is used for statistical purposes to register economic units in the national Statistical Business Register (nSBR) while the 6-digit-level is not; if Italy had introduced such a split at the lowest level, it would not have been able to produce and release statistical data separately for all seven activities considered.

A more complex example concerns NACE Rev. 2 class 26.52 "Manufacture of watches and clocks" where ATECO sub-category 26.52.00 *Fabbricazione di orologi* corresponds to five NOGA types:

265201 Manufacture and assembly of watches and clocks

265202 Manufacture and assembly of large clocks

265203 Manufacture and assembly of watch movements

265204 Manufacture of clock and watch fittings

265205 Manufacture of other watch components

4.3 Many-to-one (m:1) links

The correspondence table also includes 698 **links of many-to-one (m:1) type**, where a NOGA type corresponds to at least two ATECO sub-categories.

In three cases, a cluster contains more than 10 ATECO sub-categories linked to only one NOGA type (Table 4.4). As expected, this type of link is the most common, given that the ATECO classification comprises 1,241 codes, while the NOGA classification includes 794 codes.

Table 4.4 – Many-to-one links between ATECO 2022 and NOGA 2008

No. of ATECO sub-categories	No. of NOGA types	No. of clusters	No. of links
2	1	91	182
3	1	43	129
4	1	25	100
5	1	10	50
6	1	10	60
7	1	8	56
8	1	2	16
9	1	6	54
14	1	1	14
15	1	1	15
22	1	1	22
		198	**698**

Source: Author's processing

An example of this group of links is NACE Rev. 2 class 10.83 "Processing of tea and coffee" that has been split into two ATECO sub-categories: 10.83.01 *Lavorazione del caffè* (Processing of coffee) and 10.83.02 *Lavorazione del tè e di altri preparati per infusi* (Processing of tea and other infusions). They are both linked to NOGA 108300 *Lavorazione del tè e del caffè* (Processing of tea and coffee).

A similar case concerns ATECO sub-categories 23.52.10 "Manufacture of lime" (in Italian *Produzione di calce*) and 23.52.20 "Manufacture of plaster" (in Italian *Produzione di gesso*), which are both linked to NOGA 235200 "Manufacture of lime and plaster" (in Italian *Produzione di calce e gesso*).

A more complex case concerns NACE Rev. 2 class 33.11 corresponding exactly to NOGA 331100 "Repair of fabricated metal products" (in Italian *Riparazione e manutenzione di prodotti in metallo*) and to 8 ATECO sub-categories:

33.11.01 *Riparazione e manutenzione di stampi, portastampi, sagome, forme per macchine*
33.11.02 *Riparazione e manutenzione di utensileria ad azionamento manuale*
33.11.03 *Riparazione e manutenzione di armi, sistemi d'arma e munizioni*
33.11.04 *Riparazione e manutenzione di casseforti, forzieri, porte metalliche blindate*
33.11.05 *Riparazione e manutenzione di armi bianche*
33.11.06 *Riparazione e manutenzione di container*
33.11.07 *Riparazione e manutenzione di carrelli per la spesa*
33.11.09 *Riparazione e manutenzione di altri prodotti in metall*o

4.4 Many-to-many (m:n) links

Finally, 185 **links of many-to-many (m:n) type** have been identified.

An example of this type of link concerns the operation of dairies and cheese-making. At the Italian level, these activities are classified under two ATECO codes: 10.51.10 *Trattamento igienico del latte* and 10.51.20 *Produzione dei derivati del latte.* The same activity is instead split into three different codes in the NOGA: 105101 Manufacture of fresh dairy products, 105102 Manufacture of cheese and 105103 Other milk processing. The derived correspondence table includes four links:

Total ATECO 10.51.10 ↔ *Partial* NOGA 105101

Partial ATECO 10.51.20 ↔ *Partial* NOGA 105101

Partial ATECO 10.51.20 ↔ *Total* NOGA 105102

Partial ATECO 10.51.20 ↔ *Total* NOGA 105103

The manufacture of fresh liquid milk, for example, is classified in ATECO 10.51.10 and NOGA 105101, while the manufacture of yoghurt is classified within ATECO 10.51.20 but in NOGA 105101. The manufacture of cheese is classified in NOGA 105102, while the manufacture of whey is classified in NOGA 105103; in the ATECO classification, both the manufacture of cheese and whey are classified in the same sub-category, ATECO 10.51.20.

Another example concerns NACE class 96.01 "Washing and (dry-)cleaning of textile and fur products". The Swiss breakdown for this class is:

960101 Washing of textiles

960102 Dry-cleaning

The Italian breakdown is instead based on the type of customers served (industrial or commercial clients) and makes a distinction between self-service and non-self-service activities for commercial clients.

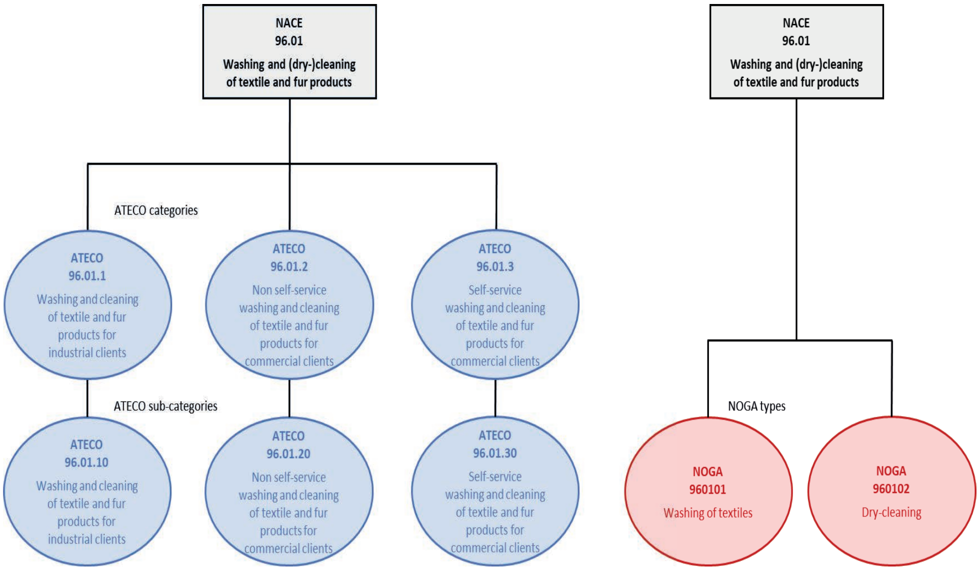
96.01.10 *Attività delle lavanderie industriali* (Washing and cleaning of textile and fur products for industrial clients)

96.01.20 *Attività di lavanderie, tintorie tradizionali* (Non-self-service washing and cleaning of textile and fur products for commercial clients)

96.01.30 *Attività di lavanderie self-service* (Self-service washing and cleaning of textile and fur products for commercial clients)

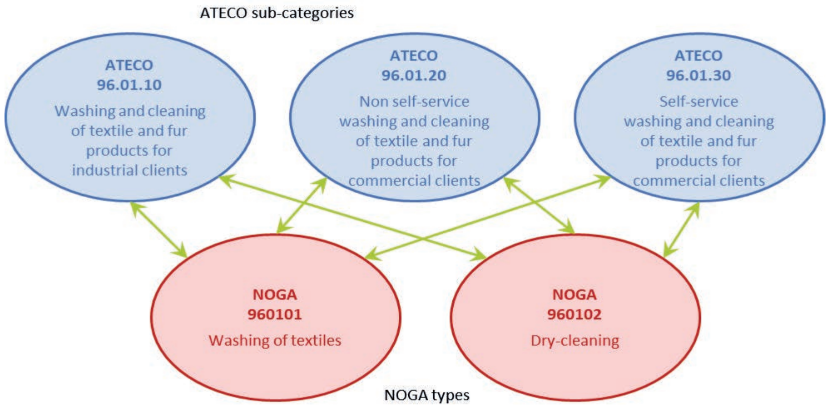
Considering all the above, the complete correspondences within NACE Rev. 2 class 96.01 between the two classifications consist of six links (Figures 4.3 and 4.4), all characterised by partial coverage from both ATECO and NOGA sides.

Figure 4.3 - ATECO 2022 and NOGA 2008 national breakdown within NACE Rev. 2 class 96.01



Source: Author’s processing

Figure 4.4 - ATECO 2022 and NOGA 2008 national many-to-many links within NACE Rev. 2 class 96.01



Source: Author's processing

4.5 Inconsistent links concerning the NACE

Regardless of the type of links created, three **critical cases** have also been identified where the correspondence at the NACE class level is not completely satisfied. In such situations, the interpretation of the same classification seems to be different in the two countries; another reason lies in the way some economic activities are organised in the different countries. Nevertheless, such a type of inconsistency should not be introduced in national classifications, thus coordination at the European level needs to be fostered, encouraging countries to share their views about classification decisions taken at the national level regularly.

The first case concerns masonry activities that are not mentioned in the NACE Rev. 2. Such activities have instead been mentioned in the ATECO classification within NACE class 43.39, but in the NOGA classification within NACE class 4399.

ATECO 43.39.01 *Attività non specializzate di lavori edili (muratori)*

NOGA 439903 Masonry (in Italian *Lavori di muratura*)

The second case, contained in a many-to-one type cluster, is about financial activities undertaken by the Post that have been included in ATECO code 66.19.40 within NACE Rev. 2 class 66.19 but in NOGA code 641912 with the following note of inclusion: “activities of Postfinance”.

ATECO 66.19.40 *Attività di Bancoposta*

NOGA 641912 Other monetary intermediation n.e.c. (in Italian A*ltre intermediazioni monetarie n.c.a.*)

The last case concerns the activities of civil protection that have been classified in ATECO code 84.25.20 within NACE Rev. 2 class 84.25 but in NOGA code 842202, generating a one-to-one link.

ATECO 84.25.20 *Attività di protezione civile*

NOGA 842202 Civil defence (in Italian *Protezione civile*)

5. Discussion and conclusions

Research in the field of statistical classifications is still scarce as expert statisticians having several cross-skills are needed to maintain, update and revise standard statistical classifications. Such classifications are essential tools in the production of statistics and are also used to support policy decisions.

The existence of an integrated system of international statistical economic classifications, developed mainly under the auspices of the United Nations Statistical Division, is the way framework statistical classifications of economic activities, products and goods, which differ in terms of family and geographical reference, are linked; thus, allowing data to be compared even if produced according to different classification schemes. Indeed, they are used to ensure coherence and especially comparability across countries, that is, between data produced in different geographical areas and statistical domains.

The system includes world, regional and national economic classifications; world classifications are linked to regional classifications, while regional classifications are linked to national classifications. As we know, no recommendations are provided to check consistency between at least two national classifications derived from the same regional (*e.g.* European) reference classifications.

The research work is intended to explore direct links between the Italian ATECO and the Swiss NOGA classifications of economic activities, both derived from the European NACE classification, to compare the two schemes, highlighting best practices and different solutions to derive national versions of the NACE. In this respect, an extension to national versions of the NACE is provided to the integrated system of statistical economic classifications.

The availability of the ATECO and the NOGA classifications in the same language (Italian) has been a strength for the whole research work. The output produced is a correspondence table between ATECO sub-categories and NOGA types, consisting of 1,397 pairwise links. The majority of the links have been derived automatically, exploiting the coding system used by the two classifications, while the remaining links, the most complex ones, have been created by analysing explanatory notes of the two classifications.

The output is quite reliable, as it has been produced by experts in the ATECO classifications. However, there is room for improvement, especially about non-automatic links. This is because the understanding and interpretation of the three classifications - NACE, ATECO, and NOGA - may vary among their respective custodians. For what we are interested in, the exploration of pairwise linking between national classifications deriving from the same common international or regional reference classification, for instance, the NACE, may have a lot of advantages. More specifically, it may serve to satisfy the following aims.

- To look for the relevance of an economic activity at European level (*e.g.* during the process of revision of the reference European classification). This goal may be reached by counting how many Member States have the same national sub-classes for a given economic activity. A good example is provided by residual classes created at European level where different activities are classified; if a certain number of Member States have defined the same national sub-class to extract one of these miscellaneous activities, it may mean that there exists a certain interest and informative need to promote it at a higher level, that of class, and thus, to create a new class in the European classification.
- To investigate the degree of consistency between national classifications directly derived from the NACE. When Member States set up national codes for economic activities that are not explicitly mentioned in the explanatory notes of a NACE class, they may generate consistency errors if their interpretation of the contents of the same European class is different. Such a circumstance may generate significant mismatches in the whole integrated system of economic classifications and thus in the data produced.
- To be used as best practices for those Member States who decide to develop a national version of the reference European classification or for all those Member States who want to revise theirs. A regular process of comparison between national classifications derived by the same reference classification may also be useful to define European guidelines to support Member States in deriving national versions.

- To provide information to European and national stakeholders on the availability of comparable national data referred to some countries that would result in being more detailed than those available at European level.
- In the long period, those Member States whose productive systems are similar, might also consider aligning their national classifications, at least for some classes whose relevance is recognised in the different countries, to release more comparable data and look for harmonisation.

Despite of the advantages that may be derived, the development and maintenance of *ad hoc* pairwise correspondences are not a common practice especially because it is time-consuming and requires the involvement of experts in the field of statistical classifications who are rare. It also may necessitate the knowledge of different languages, since those national classifications are not always available in English or in the same language.

Indeed, once a pairwise correspondence table between two national classifications (*e.g.* national versions of the European NACE) is set up for the first time, the activities to maintain and update such a correspondence table are meant to be sufficient easy as it means systematically applying correspondences between the new NACE and its new national versions. But it also means introducing manual checks based on a deep study of explanatory notes to solve the trickiest links.

The results of this research work should encourage experts in the field of statistical classifications to explore pairwise links between national versions of the same European classifications to check if they are really consistent within the whole integrated system of international statistical economic classifications. In addition, such correspondence tables may be used at European level when a country encounters a classification problem that has already been dealt and solved within a national version.

To achieve this goal, national versions should also be available in English. When two countries apply the same sub-classes, their title in English should be the same to facilitate the understanding process of the users.

At the moment of writing, the revision process of part of the integrated system of statistical economic classifications activated at international and European level is reaching an end. More specifically, the new NACE Rev. 2.1 will enter into force starting on 1 January 2025; at the same time, existing national versions will have been revised as well and adapted to the new structure of the NACE. From then on, a process of regularly updating the classifications of economic activities is foreseen *e.g.* the creation of new index entries at European level. Such a decision has a significant impact on national versions which are in most cases already more detailed than the NACE implying a regular process to integrate European index entries in national versions. In this context ensuring consistency between national versions would acquire even more importance.

Appendix - Correspondence table between ATECO and NOGA

Coverage ATECO	ATECO 2022	Coverage NOGA	NOGA 2008	Link type
Total	01.11.10	Partial	011100	m:1
Total	01.11.20	Partial	011100	m:1
Total	01.11.30	Partial	011100	m:1
Total	01.11.40	Partial	011100	m:1
Total	01.12.00	Total	011200	1:1
Total	01.13.10	Partial	011300	m:1
Total	01.13.21	Partial	011300	m:1
Total	01.13.29	Partial	011300	m:1
Total	01.13.30	Partial	011300	m:1
Total	01.13.40	Partial	011300	m:1
Total	01.14.00	Total	011400	1:1
Total	01.15.00	Total	011500	1:1
Total	01.16.00	Total	011600	1:1
Total	01.19.10	Partial	011900	m:1
Total	01.19.21	Partial	011900	m:1
Total	01.19.29	Partial	011900	m:1
Total	01.19.90	Partial	011900	m:1
Partial	01.21.00	Total	012101	1:n
Partial	01.21.00	Total	012102	1:n
Total	01.22.00	Total	012200	1:1
Total	01.23.00	Total	012300	1:1
Total	01.24.00	Total	012400	1:1
Total	01.25.00	Total	012500	1:1
Total	01.26.00	Total	012600	1:1
Total	01.27.00	Total	012700	1:1
Total	01.28.00	Total	012800	1:1
Total	01.29.00	Total	012900	1:1
Total	01.30.00	Total	013000	1:1
Total	01.41.00	Total	014100	1:1
Total	01.42.00	Total	014200	1:1
Total	01.43.00	Total	014300	1:1
Total	01.44.00	Total	014400	1:1
Total	01.45.00	Total	014500	1:1
Total	01.46.00	Total	014600	1:1
Total	01.47.00	Total	014700	1:1
Total	01.49.10	Partial	014900	m:1
Total	01.49.20	Partial	014900	m:1
Total	01.49.30	Partial	014900	m:1
Total	01.49.40	Partial	014900	m:1
Total	01.49.90	Partial	014900	m:1
Total	01.50.00	Total	015000	1:1
Total	01.61.00	Total	016100	1:1

Coverage ATECO	ATECO 2022	Coverage NOGA	NOGA 2008	Link type
Total	01.62.01	Partial	016200	m:1
Total	01.62.09	Partial	016200	m:1
Total	01.63.00	Total	016300	1:1
Total	01.64.01	Partial	016400	m:1
Total	01.64.09	Partial	016400	m:1
Total	01.70.00	Total	017000	1:1
Total	02.10.00	Total	021000	1:1
Total	02.20.00	Total	022000	1:1
Total	02.30.00	Total	023000	1:1
Total	02.40.00	Total	024000	1:1
Total	03.11.00	Total	031100	1:1
Total	03.12.00	Total	031200	1:1
Total	03.21.00	Total	032100	1:1
Total	03.22.00	Total	032200	1:1
Total	05.10.00	Total	051000	1:1
Total	05.20.00	Total	052000	1:1
Total	06.10.00	Total	061000	1:1
Total	06.20.00	Total	062000	1:1
Total	07.10.00	Total	071000	1:1
Total	07.21.00	Total	072100	1:1
Total	07.29.00	Total	072900	1:1
Total	08.11.00	Total	081100	1:1
Total	08.12.00	Total	081200	1:1
Total	08.91.00	Total	089100	1:1
Total	08.92.00	Total	089200	1:1
Total	08.93.00	Total	089300	1:1
Total	08.99.01	Partial	089900	m:1
Total	08.99.09	Partial	089900	m:1
Total	09.10.00	Total	091000	1:1
Total	09.90.01	Partial	099000	m:1
Total	09.90.09	Partial	099000	m:1
Total	10.11.00	Total	101100	1:1
Total	10.12.00	Total	101200	1:1
Total	10.13.00	Total	101300	1:1
Total	10.20.00	Total	102000	1:1
Total	10.31.00	Total	103100	1:1
Total	10.32.00	Total	103200	1:1
Total	10.39.00	Total	103900	1:1
Total	10.41.10	Partial	104100	m:1
Total	10.41.20	Partial	104100	m:1
Total	10.41.30	Partial	104100	m:1
Total	10.42.00	Total	104200	1:1
Total	10.51.10	Partial	105101	m:n
Partial	10.51.20	Partial	105101	m:n

Coverage ATECO	ATECO 2022	Coverage NOGA	NOGA 2008	Link type
Partial	10.51.20	Total	105102	m:n
Partial	10.51.20	Total	105103	m:n
Total	10.52.00	Total	105200	1:1
Total	10.61.10	Partial	106100	m:1
Total	10.61.20	Partial	106100	m:1
Total	10.61.30	Partial	106100	m:1
Total	10.61.40	Partial	106100	m:1
Total	10.62.00	Total	106200	1:1
Total	10.71.10	Partial	107100	m:1
Total	10.71.20	Partial	107100	m:1
Total	10.72.00	Total	107200	1:1
Total	10.73.00	Total	107300	1:1
Total	10.81.00	Total	108100	1:1
Partial	10.82.00	Total	108201	1:n
Partial	10.82.00	Total	108202	1:n
Total	10.83.01	Partial	108300	m:1
Total	10.83.02	Partial	108300	m:1
Total	10.84.00	Total	108400	1:1
Total	10.85.01	Partial	108500	m:1
Total	10.85.02	Partial	108500	m:1
Total	10.85.03	Partial	108500	m:1
Total	10.85.04	Partial	108500	m:1
Total	10.85.05	Partial	108500	m:1
Total	10.85.09	Partial	108500	m:1
Total	10.86.00	Total	108600	1:1
Total	10.89.01	Partial	108900	m:1
Total	10.89.09	Partial	108900	m:1
Total	10.91.00	Total	109100	1:1
Total	10.92.00	Total	109200	1:1
Total	11.01.00	Total	110100	1:1
Total	11.02.10	Partial	110200	m:1
Total	11.02.20	Partial	110200	m:1
Total	11.03.00	Total	110300	1:1
Total	11.04.00	Total	110400	1:1
Total	11.05.00	Total	110500	1:1
Total	11.06.00	Total	110600	1:1
Total	11.07.00	Total	110700	1:1
Total	12.00.00	Total	120000	1:1
Partial	13.10.00	Total	131001	1:n
Partial	13.10.00	Total	131002	1:n
Partial	13.10.00	Total	131003	1:n
Partial	13.10.00	Total	131004	1:n
Partial	13.20.00	Total	132001	1:n
Partial	13.20.00	Total	132002	1:n

Coverage ATECO	ATECO 2022	Coverage NOGA	NOGA 2008	Link type
Partial	13.20.00	Total	132003	1:n
Total	13.30.00	Total	133000	1:1
Total	13.91.00	Total	139100	1:1
Partial	13.92.10	Partial	139201	m:n
Partial	13.92.10	Partial	139202	m:n
Partial	13.92.20	Total	139202	m:n
Partial	13.92.20	Total	139203	m:n
Total	13.93.00	Total	139300	1:1
Total	13.94.00	Total	139400	1:1
Total	13.95.00	Total	139500	1:1
Total	13.96.10	Partial	139600	m:1
Total	13.96.20	Partial	139600	m:1
Total	13.99.10	Total	139901	1:1
Total	13.99.20	Total	139902	1:1
Total	13.99.90	Total	139903	1:1
Total	14.11.00	Total	141100	1:1
Total	14.12.00	Total	141200	1:1
Partial	14.13.10	Partial	141301	m:n
Partial	14.13.10	Partial	141302	m:n
Partial	14.13.10	Partial	141303	m:n
Partial	14.13.20	Partial	141301	m:n
Partial	14.13.20	Partial	141302	m:n
Partial	14.13.20	Partial	141303	m:n
Partial	14.14.00	Total	141401	1:n
Partial	14.14.00	Total	141402	1:n
Partial	14.14.00	Total	141403	1:n
Total	14.19.10	Partial	141900	m:1
Total	14.19.21	Partial	141900	m:1
Total	14.19.29	Partial	141900	m:1
Total	14.20.00	Total	142000	1:1
Total	14.31.00	Total	143100	1:1
Total	14.39.00	Total	143900	1:1
Total	15.11.00	Total	151100	1:1
Total	15.12.01	Partial	151200	m:1
Total	15.12.09	Partial	151200	m:1
Total	15.20.10	Partial	152000	m:1
Total	15.20.20	Partial	152000	m:1
Partial	16.10.00	Total	161001	1:n
Partial	16.10.00	Total	161002	1:n
Partial	16.10.00	Total	161003	1:n
Total	16.21.00	Total	162100	1:1
Total	16.22.00	Total	162200	1:1
Total	16.23.10	Partial	162301	m:n
Total	16.23.21	Partial	162303	m:n

Coverage ATECO	ATECO 2022	Coverage NOGA	NOGA 2008	Link type
Partial	16.23.22	Partial	162301	m:n
Partial	16.23.22	Total	162302	m:n
Partial	16.23.22	Partial	162303	m:n
Total	16.24.00	Total	162400	1:1
Total	16.29.11	Partial	162900	m:1
Total	16.29.12	Partial	162900	m:1
Total	16.29.19	Partial	162900	m:1
Total	16.29.20	Partial	162900	m:1
Total	16.29.30	Partial	162900	m:1
Total	16.29.40	Partial	162900	m:1
Total	17.11.00	Total	171100	1:1
Total	17.12.00	Total	171200	1:1
Total	17.21.00	Total	172100	1:1
Total	17.22.00	Total	172200	1:1
Total	17.23.01	Partial	172300	m:1
Total	17.23.09	Partial	172300	m:1
Total	17.24.00	Total	172400	1:1
Total	17.29.00	Total	172900	1:1
Total	18.11.00	Total	181100	1:1
Partial	18.12.00	Total	181201	1:n
Partial	18.12.00	Total	181202	1:n
Partial	18.12.00	Total	181203	1:n
Partial	18.12.00	Total	181204	1:n
Partial	18.13.00	Total	181301	1:n
Partial	18.13.00	Total	181302	1:n
Total	18.14.00	Total	181400	1:1
Total	18.20.00	Total	182000	1:1
Total	19.10.01	Partial	191000	m:1
Total	19.10.09	Partial	191000	m:1
Total	19.20.10	Partial	192000	m:1
Total	19.20.20	Partial	192000	m:1
Total	19.20.30	Partial	192000	m:1
Total	19.20.40	Partial	192000	m:1
Total	19.20.90	Partial	192000	m:1
Total	20.11.00	Total	201100	1:1
Total	20.12.00	Total	201200	1:1
Total	20.13.01	Partial	201300	m:1
Total	20.13.09	Partial	201300	m:1
Total	20.14.01	Partial	201400	m:1
Total	20.14.09	Partial	201400	m:1
Total	20.15.00	Total	201500	1:1
Total	20.16.00	Total	201600	1:1
Total	20.17.00	Total	201700	1:1
Total	20.20.00	Total	202000	1:1

Coverage ATECO	**ATECO 2022**	**Coverage NOGA**	**NOGA 2008**	**Link type**
Total	20.30.00	Total	203000	1:1
Total	20.41.10	Partial	204100	m:1
Total	20.41.20	Partial	204100	m:1
Total	20.42.00	Total	204200	1:1
Total	20.51.01	Partial	205100	m:1
Total	20.51.02	Partial	205100	m:1
Total	20.52.00	Total	205200	1:1
Total	20.53.00	Total	205300	1:1
Total	20.59.10	Partial	205900	m:1
Total	20.59.20	Partial	205900	m:1
Total	20.59.30	Partial	205900	m:1
Total	20.59.40	Partial	205900	m:1
Total	20.59.50	Partial	205900	m:1
Total	20.59.60	Partial	205900	m:1
Total	20.59.70	Partial	205900	m:1
Total	20.59.90	Partial	205900	m:1
Total	20.60.00	Total	206000	1:1
Total	21.10.00	Total	211000	1:1
Total	21.20.01	Partial	212000	m:1
Total	21.20.09	Partial	212000	m:1
Total	22.11.10	Partial	221100	m:1
Total	22.11.20	Partial	221100	m:1
Total	22.19.01	Partial	221900	m:1
Total	22.19.09	Partial	221900	m:1
Total	22.21.00	Total	222100	1:1
Total	22.22.00	Total	222200	1:1
Total	22.23.01	Partial	222300	m:1
Total	22.23.02	Partial	222300	m:1
Total	22.23.09	Partial	222300	m:1
Total	22.29.01	Partial	222900	m:1
Total	22.29.02	Partial	222900	m:1
Total	22.29.09	Partial	222900	m:1
Total	23.11.00	Total	231100	1:1
Total	23.12.00	Total	231200	1:1
Total	23.13.00	Total	231300	1:1
Total	23.14.00	Total	231400	1:1
Total	23.19.10	Partial	231900	m:1
Total	23.19.20	Partial	231900	m:1
Total	23.19.90	Partial	231900	m:1
Total	23.20.00	Total	232000	1:1
Total	23.31.00	Total	233100	1:1
Total	23.32.00	Total	233200	1:1
Total	23.41.00	Total	234100	1:1
Total	23.42.00	Total	234200	1:1

Coverage ATECO	**ATECO 2022**	**Coverage NOGA**	**NOGA 2008**	**Link type**
Total	23.43.00	Total	234300	1:1
Total	23.44.00	Total	234400	1:1
Total	23.49.00	Total	234900	1:1
Total	23.51.00	Total	235100	1:1
Total	23.52.10	Partial	235200	m:1
Total	23.52.20	Partial	235200	m:1
Total	23.61.00	Total	236100	1:1
Total	23.62.00	Total	236200	1:1
Total	23.63.00	Total	236300	1:1
Total	23.64.00	Total	236400	1:1
Total	23.65.00	Total	236500	1:1
Total	23.69.00	Total	236900	1:1
Total	23.70.10	Partial	237000	m:1
Total	23.70.20	Partial	237000	m:1
Total	23.70.30	Partial	237000	m:1
Total	23.91.00	Total	239100	1:1
Partial	23.99.00	Total	239901	1:n
Partial	23.99.00	Total	239902	1:n
Total	24.10.00	Total	241000	1:1
Total	24.20.10	Partial	242000	m:1
Total	24.20.20	Partial	242000	m:1
Total	24.31.00	Total	243100	1:1
Total	24.32.00	Total	243200	1:1
Total	24.33.01	Partial	243300	m:1
Total	24.33.02	Partial	243300	m:1
Total	24.33.03	Partial	243300	m:1
Total	24.34.00	Total	243400	1:1
Total	24.41.00	Total	244100	1:1
Total	24.42.00	Total	244200	1:1
Total	24.43.00	Total	244300	1:1
Total	24.44.00	Total	244400	1:1
Total	24.45.00	Total	244500	1:1
Total	24.46.00	Total	244600	1:1
Total	24.51.00	Total	245100	1:1
Total	24.52.00	Total	245200	1:1
Total	24.53.00	Total	245300	1:1
Total	24.54.00	Total	245400	1:1
Total	25.11.00	Total	251100	1:1
Total	25.12.10	Partial	251200	m:1
Total	25.12.20	Partial	251200	m:1
Total	25.21.00	Total	252100	1:1
Total	25.29.00	Total	252900	1:1
Total	25.30.00	Total	253000	1:1
Total	25.40.00	Total	254000	1:1

Coverage ATECO	**ATECO 2022**	**Coverage NOGA**	**NOGA 2008**	**Link type**
Total	25.50.00	Total	255000	1:1
Total	25.61.00	Total	256100	1:1
Partial	25.62.00	Total	256201	1:n
Partial	25.62.00	Total	256202	1:n
Partial	25.62.00	Total	256203	1:n
Total	25.71.00	Total	257100	1:1
Total	25.72.00	Total	257200	1:1
Total	25.73.11	Partial	257300	m:1
Total	25.73.12	Partial	257300	m:1
Total	25.73.20	Partial	257300	m:1
Total	25.91.00	Total	259100	1:1
Total	25.92.00	Total	259200	1:1
Total	25.93.10	Partial	259300	m:1
Total	25.93.20	Partial	259300	m:1
Total	25.93.30	Partial	259300	m:1
Total	25.94.00	Total	259400	1:1
Total	25.99.11	Partial	259900	m:1
Total	25.99.19	Partial	259900	m:1
Total	25.99.20	Partial	259900	m:1
Total	25.99.30	Partial	259900	m:1
Total	25.99.91	Partial	259900	m:1
Total	25.99.99	Partial	259900	m:1
Total	26.11.01	Partial	261100	m:1
Total	26.11.09	Partial	261100	m:1
Total	26.12.00	Total	261200	1:1
Total	26.20.00	Total	262000	1:1
Total	26.30.10	Partial	263000	m:1
Total	26.30.21	Partial	263000	m:1
Total	26.30.29	Partial	263000	m:1
Total	26.40.01	Partial	264000	m:1
Total	26.40.02	Partial	264000	m:1
Total	26.51.10	Partial	265100	m:1
Total	26.51.21	Partial	265100	m:1
Total	26.51.29	Partial	265100	m:1
Partial	26.52.00	Total	265201	1:n
Partial	26.52.00	Total	265202	1:n
Partial	26.52.00	Total	265203	1:n
Partial	26.52.00	Total	265204	1:n
Partial	26.52.00	Total	265205	1:n
Total	26.60.01	Partial	266000	m:1
Total	26.60.02	Partial	266000	m:1
Total	26.60.09	Partial	266000	m:1
Total	26.70.11	Partial	267000	m:1
Total	26.70.12	Partial	267000	m:1

Coverage ATECO	ATECO 2022	Coverage NOGA	NOGA 2008	Link type
Total	26.70.20	Partial	267000	m:1
Total	26.80.00	Total	268000	1:1
Total	27.11.00	Total	271100	1:1
Total	27.12.00	Total	271200	1:1
Total	27.20.00	Total	272000	1:1
Total	27.31.01	Partial	273100	m:1
Total	27.31.02	Partial	273100	m:1
Total	27.32.00	Total	273200	1:1
Total	27.33.01	Partial	273300	m:1
Total	27.33.09	Partial	273300	m:1
Total	27.40.01	Partial	274000	m:1
Total	27.40.02	Partial	274000	m:1
Total	27.40.09	Partial	274000	m:1
Total	27.51.00	Total	275100	1:1
Total	27.52.00	Total	275200	1:1
Total	27.90.01	Partial	279000	m:1
Total	27.90.02	Partial	279000	m:1
Total	27.90.03	Partial	279000	m:1
Total	27.90.09	Partial	279000	m:1
Total	28.11.11	Partial	281100	m:1
Total	28.11.12	Partial	281100	m:1
Total	28.11.20	Partial	281100	m:1
Total	28.12.00	Total	281200	1:1
Total	28.13.00	Total	281300	1:1
Total	28.14.00	Total	281400	1:1
Total	28.15.10	Partial	281500	m:1
Total	28.15.20	Partial	281500	m:1
Total	28.21.10	Partial	282100	m:1
Total	28.21.21	Partial	282100	m:1
Total	28.21.29	Partial	282100	m:1
Total	28.22.01	Partial	282200	m:1
Total	28.22.02	Partial	282200	m:1
Total	28.22.03	Partial	282200	m:1
Total	28.22.09	Partial	282200	m:1
Total	28.23.01	Partial	282300	m:1
Total	28.23.09	Partial	282300	m:1
Total	28.24.00	Total	282400	1:1
Total	28.25.00	Total	282500	1:1
Total	28.29.10	Partial	282900	m:1
Total	28.29.20	Partial	282900	m:1
Total	28.29.30	Partial	282900	m:1
Total	28.29.91	Partial	282900	m:1
Total	28.29.92	Partial	282900	m:1
Total	28.29.93	Partial	282900	m:1

Coverage ATECO	ATECO 2022	Coverage NOGA	NOGA 2008	Link type
Total	28.29.99	Partial	282900	m:1
Total	28.30.10	Partial	283000	m:1
Total	28.30.90	Partial	283000	m:1
Total	28.41.00	Total	284100	1:1
Total	28.49.01	Partial	284900	m:1
Total	28.49.09	Partial	284900	m:1
Total	28.91.00	Total	289100	1:1
Total	28.92.01	Partial	289200	m:1
Total	28.92.09	Partial	289200	m:1
Total	28.93.00	Total	289300	1:1
Total	28.94.10	Partial	289400	m:1
Total	28.94.20	Partial	289400	m:1
Total	28.94.30	Partial	289400	m:1
Total	28.95.00	Total	289500	1:1
Total	28.96.00	Total	289600	1:1
Total	28.99.10	Total	289901	1:1
Total	28.99.20	Partial	289902	m:1
Total	28.99.30	Partial	289902	m:1
Total	28.99.91	Partial	289902	m:1
Total	28.99.92	Partial	289902	m:1
Total	28.99.93	Partial	289902	m:1
Total	28.99.99	Partial	289902	m:1
Total	29.10.00	Total	291000	1:1
Total	29.20.00	Total	292000	1:1
Total	29.31.00	Total	293100	1:1
Total	29.32.01	Partial	293200	m:1
Total	29.32.09	Partial	293200	m:1
Total	30.11.01	Partial	301100	m:1
Total	30.11.02	Partial	301100	m:1
Total	30.12.00	Total	301200	1:1
Total	30.20.01	Partial	302000	m:1
Total	30.20.02	Partial	302000	m:1
Total	30.30.01	Partial	303000	m:1
Total	30.30.02	Partial	303000	m:1
Total	30.30.09	Partial	303000	m:1
Total	30.40.00	Total	304000	1:1
Total	30.91.11	Partial	309100	m:1
Total	30.91.12	Partial	309100	m:1
Total	30.91.20	Partial	309100	m:1
Total	30.92.10	Partial	309201	m:1
Total	30.92.20	Partial	309201	m:1
Total	30.92.30	Total	309202	1:1
Total	30.92.40	Partial	309201	m:1
Total	30.99.00	Total	309900	1:1

Coverage ATECO	ATECO 2022	Coverage NOGA	NOGA 2008	Link type
Total	31.01.10	Partial	310100	m:1
Total	31.01.21	Partial	310100	m:1
Total	31.01.22	Partial	310100	m:1
Total	31.02.00	Total	310200	1:1
Total	31.03.00	Total	310300	1:1
Total	31.09.10	Partial	310900	m:1
Total	31.09.20	Partial	310900	m:1
Total	31.09.30	Partial	310900	m:1
Total	31.09.40	Partial	310900	m:1
Total	31.09.50	Partial	310900	m:1
Total	31.09.90	Partial	310900	m:1
Total	32.11.00	Total	321100	1:1
Total	32.12.10	Total	321202	1:1
Total	32.12.20	Total	321201	1:1
Total	32.13.01	Partial	321300	m:1
Total	32.13.09	Partial	321300	m:1
Total	32.20.00	Total	322000	1:1
Total	32.30.00	Total	323000	1:1
Total	32.40.10	Partial	324000	m:1
Total	32.40.20	Partial	324000	m:1
Total	32.50.11	Partial	325001	m:1
Total	32.50.12	Partial	325001	m:1
Total	32.50.13	Partial	325001	m:1
Total	32.50.14	Partial	325001	m:1
Total	32.50.20	Total	325003	1:1
Total	32.50.30	Total	325002	1:1
Total	32.50.40	Partial	325004	m:1
Total	32.50.50	Partial	325004	m:1
Total	32.91.00	Total	329100	1:1
Total	32.99.11	Partial	329900	m:1
Total	32.99.12	Partial	329900	m:1
Total	32.99.13	Partial	329900	m:1
Total	32.99.14	Partial	329900	m:1
Total	32.99.19	Partial	329900	m:1
Total	32.99.20	Partial	329900	m:1
Total	32.99.30	Partial	329900	m:1
Total	32.99.40	Partial	329900	m:1
Total	32.99.90	Partial	329900	m:1
Total	33.11.01	Partial	331100	m:1
Total	33.11.02	Partial	331100	m:1
Total	33.11.03	Partial	331100	m:1
Total	33.11.04	Partial	331100	m:1
Total	33.11.05	Partial	331100	m:1
Total	33.11.06	Partial	331100	m:1

Coverage ATECO	**ATECO 2022**	**Coverage NOGA**	**NOGA 2008**	**Link type**
Total	33.11.07	Partial	331100	m:1
Total	33.11.09	Partial	331100	m:1
Total	33.12.10	Partial	331200	m:1
Total	33.12.20	Partial	331200	m:1
Total	33.12.30	Partial	331200	m:1
Total	33.12.40	Partial	331200	m:1
Total	33.12.51	Partial	331200	m:1
Total	33.12.52	Partial	331200	m:1
Total	33.12.53	Partial	331200	m:1
Total	33.12.54	Partial	331200	m:1
Total	33.12.55	Partial	331200	m:1
Total	33.12.59	Partial	331200	m:1
Total	33.12.60	Partial	331200	m:1
Total	33.12.70	Partial	331200	m:1
Total	33.12.91	Partial	331200	m:1
Total	33.12.92	Partial	331200	m:1
Total	33.12.99	Partial	331200	m:1
Total	33.13.01	Partial	331300	m:1
Total	33.13.03	Partial	331300	m:1
Total	33.13.04	Partial	331300	m:1
Total	33.13.09	Partial	331300	m:1
Total	33.14.00	Total	331400	1:1
Total	33.15.00	Total	331500	1:1
Total	33.16.00	Total	331600	1:1
Total	33.17.00	Total	331700	1:1
Total	33.19.01	Partial	331900	m:1
Total	33.19.02	Partial	331900	m:1
Total	33.19.03	Partial	331900	m:1
Total	33.19.04	Partial	331900	m:1
Total	33.19.09	Partial	331900	m:1
Total	33.20.01	Partial	332000	m:1
Total	33.20.02	Partial	332000	m:1
Total	33.20.03	Partial	332000	m:1
Total	33.20.04	Partial	332000	m:1
Total	33.20.05	Partial	332000	m:1
Total	33.20.06	Partial	332000	m:1
Total	33.20.07	Partial	332000	m:1
Total	33.20.08	Partial	332000	m:1
Total	33.20.09	Partial	332000	m:1
Total	35.11.00	Total	351100	1:1
Total	35.12.00	Total	351200	1:1
Total	35.13.00	Total	351300	1:1
Total	35.14.00	Total	351400	1:1
Total	35.21.00	Total	352100	1:1

Coverage ATECO	**ATECO 2022**	**Coverage NOGA**	**NOGA 2008**	**Link type**
Total	35.22.00	Total	352200	1:1
Total	35.23.00	Total	352300	1:1
Total	35.30.00	Total	353000	1:1
Total	36.00.00	Total	360000	1:1
Total	37.00.00	Total	370000	1:1
Total	38.11.00	Total	381100	1:1
Total	38.12.00	Total	381200	1:1
Total	38.21.01	Partial	382100	m:1
Total	38.21.09	Partial	382100	m:1
Total	38.22.00	Total	382200	1:1
Total	38.31.10	Partial	383100	m:1
Total	38.31.20	Partial	383100	m:1
Total	38.32.10	Partial	383200	m:1
Total	38.32.20	Partial	383200	m:1
Total	38.32.30	Partial	383200	m:1
Total	39.00.01	Partial	390000	m:1
Total	39.00.09	Partial	390000	m:1
Total	41.10.00	Total	411000	1:1
Partial	41.20.00	Total	412001	1:n
Partial	41.20.00	Total	412002	1:n
Partial	41.20.00	Total	412003	1:n
Partial	41.20.00	Total	412004	1:n
Total	42.11.00	Total	421100	1:1
Total	42.12.00	Total	421200	1:1
Total	42.13.00	Total	421300	1:1
Total	42.21.00	Total	422100	1:1
Total	42.22.00	Total	422200	1:1
Total	42.91.00	Total	429100	1:1
Total	42.99.01	Partial	429900	m:1
Total	42.99.09	Partial	429900	m:1
Total	43.11.00	Total	431100	1:1
Total	43.12.00	Total	431200	1:1
Total	43.13.00	Total	431300	1:1
Total	43.21.01	Partial	432100	m:1
Total	43.21.02	Partial	432100	m:1
Total	43.21.03	Partial	432100	m:1
Total	43.21.04	Partial	432100	m:1
Partial	43.22.01	Total	432201	m:n
Partial	43.22.01	Partial	432202	m:n
Partial	43.22.01	Total	432203	m:n
Total	43.22.02	Partial	432202	m:n
Total	43.22.03	Partial	432204	m:1
Total	43.22.04	Partial	432204	m:1
Total	43.22.05	Partial	432204	m:1

Coverage ATECO	ATECO 2022	Coverage NOGA	NOGA 2008	Link type
Total	43.29.01	Partial	432902	m:1
Total	43.29.02	Total	432901	1:1
Total	43.29.09	Partial	432902	m:1
Total	43.31.00	Total	433100	1:1
Total	43.32.01	Partial	433200	m:1
Total	43.32.02	Partial	433200	m:1
Partial	43.33.00	Total	433301	1:n
Partial	43.33.00	Total	433302	1:n
Partial	43.33.00	Total	433303	1:n
Partial	43.34.00	Total	433401	1:n
Partial	43.34.00	Total	433402	1:n
Partial	43.34.00	Total	433403	1:n
Partial	43.39.01	Partial	433900	m:n
Partial	43.39.01	Partial	439903	m:n
Partial	43.39.09	Partial	433900	m:n
Partial	43.39.09	Partial	439903	m:n
Partial	43.91.00	Total	439101	1:n
Partial	43.91.00	Total	439102	1:n
Partial	43.91.00	Total	439103	1:n
Total	43.99.01	Partial	439905	m:n
Total	43.99.02	Total	439904	1:1
Partial	43.99.09	Total	439901	m:n
Partial	43.99.09	Total	439902	m:n
Partial	43.99.09	Partial	439905	m:n
Partial	45.11.01	Partial	451101	m:n
Partial	45.11.01	Total	451102	m:n
Total	45.11.02	Partial	451101	m:n
Partial	45.19.01	Partial	451901	m:n
Partial	45.19.01	Total	451902	m:n
Total	45.19.02	Partial	451901	m:n
Total	45.20.10	Partial	452001	m:1
Total	45.20.20	Total	452002	1:1
Total	45.20.30	Partial	452001	m:1
Total	45.20.40	Partial	452001	m:1
Total	45.20.91	Partial	452001	m:1
Total	45.20.99	Partial	452001	m:1
Total	45.31.01	Partial	453100	m:1
Total	45.31.02	Partial	453100	m:1
Total	45.32.00	Total	453200	1:1
Total	45.40.11	Partial	454000	m:1
Total	45.40.12	Partial	454000	m:1
Total	45.40.21	Partial	454000	m:1
Total	45.40.22	Partial	454000	m:1
Total	45.40.30	Partial	454000	m:1

Coverage ATECO	ATECO 2022	Coverage NOGA	NOGA 2008	Link type
Total	46.11.01	Partial	461100	m:1
Total	46.11.02	Partial	461100	m:1
Total	46.11.03	Partial	461100	m:1
Total	46.11.04	Partial	461100	m:1
Total	46.11.05	Partial	461100	m:1
Total	46.11.06	Partial	461100	m:1
Total	46.11.07	Partial	461100	m:1
Total	46.12.01	Partial	461200	m:1
Total	46.12.02	Partial	461200	m:1
Total	46.12.03	Partial	461200	m:1
Total	46.12.04	Partial	461200	m:1
Total	46.12.05	Partial	461200	m:1
Total	46.12.06	Partial	461200	m:1
Total	46.12.07	Partial	461200	m:1
Total	46.13.01	Partial	461300	m:1
Total	46.13.02	Partial	461300	m:1
Total	46.13.03	Partial	461300	m:1
Total	46.13.04	Partial	461300	m:1
Total	46.13.05	Partial	461300	m:1
Total	46.14.01	Partial	461400	m:1
Total	46.14.02	Partial	461400	m:1
Total	46.14.03	Partial	461400	m:1
Total	46.14.04	Partial	461400	m:1
Total	46.14.05	Partial	461400	m:1
Total	46.14.06	Partial	461400	m:1
Total	46.14.07	Partial	461400	m:1
Total	46.15.01	Partial	461500	m:1
Total	46.15.02	Partial	461500	m:1
Total	46.15.03	Partial	461500	m:1
Total	46.15.04	Partial	461500	m:1
Total	46.15.05	Partial	461500	m:1
Total	46.15.06	Partial	461500	m:1
Total	46.15.07	Partial	461500	m:1
Total	46.16.01	Partial	461600	m:1
Total	46.16.02	Partial	461600	m:1
Total	46.16.03	Partial	461600	m:1
Total	46.16.04	Partial	461600	m:1
Total	46.16.05	Partial	461600	m:1
Total	46.16.06	Partial	461600	m:1
Total	46.16.07	Partial	461600	m:1
Total	46.16.08	Partial	461600	m:1
Total	46.16.09	Partial	461600	m:1
Total	46.17.01	Partial	461700	m:1
Total	46.17.02	Partial	461700	m:1

Coverage ATECO	ATECO 2022	Coverage NOGA	NOGA 2008	Link type
Total	46.17.03	Partial	461700	m:1
Total	46.17.04	Partial	461700	m:1
Total	46.17.05	Partial	461700	m:1
Total	46.17.06	Partial	461700	m:1
Total	46.17.07	Partial	461700	m:1
Total	46.17.08	Partial	461700	m:1
Total	46.17.09	Partial	461700	m:1
Total	46.18.11	Partial	461800	m:1
Total	46.18.12	Partial	461800	m:1
Total	46.18.13	Partial	461800	m:1
Total	46.18.14	Partial	461800	m:1
Total	46.18.21	Partial	461800	m:1
Total	46.18.22	Partial	461800	m:1
Total	46.18.23	Partial	461800	m:1
Total	46.18.24	Partial	461800	m:1
Total	46.18.31	Partial	461800	m:1
Total	46.18.32	Partial	461800	m:1
Total	46.18.33	Partial	461800	m:1
Total	46.18.34	Partial	461800	m:1
Total	46.18.35	Partial	461800	m:1
Total	46.18.91	Partial	461800	m:1
Total	46.18.92	Partial	461800	m:1
Total	46.18.93	Partial	461800	m:1
Total	46.18.94	Partial	461800	m:1
Total	46.18.95	Partial	461800	m:1
Total	46.18.96	Partial	461800	m:1
Total	46.18.97	Partial	461800	m:1
Total	46.18.98	Partial	461800	m:1
Total	46.18.99	Partial	461800	m:1
Total	46.19.01	Partial	461900	m:1
Total	46.19.02	Partial	461900	m:1
Total	46.19.03	Partial	461900	m:1
Total	46.19.04	Partial	461900	m:1
Total	46.21.10	Partial	462100	m:1
Total	46.21.21	Partial	462100	m:1
Total	46.21.22	Partial	462100	m:1
Total	46.22.00	Total	462200	1:1
Total	46.23.00	Total	462300	1:1
Total	46.24.10	Partial	462400	m:1
Total	46.24.20	Partial	462400	m:1
Total	46.31.10	Partial	463100	m:1
Total	46.31.20	Partial	463100	m:1
Total	46.32.10	Partial	463200	m:1
Total	46.32.20	Partial	463200	m:1

Coverage ATECO	ATECO 2022	Coverage NOGA	NOGA 2008	Link type
Total	46.33.10	Partial	463300	m:1
Total	46.33.20	Partial	463300	m:1
Total	46.34.10	Total	463401	1:1
Total	46.34.20	Total	463402	1:1
Total	46.35.00	Total	463500	1:1
Total	46.36.00	Total	463600	1:1
Total	46.37.01	Partial	463700	m:1
Total	46.37.02	Partial	463700	m:1
Total	46.38.10	Partial	463800	m:1
Total	46.38.20	Partial	463800	m:1
Total	46.38.30	Partial	463800	m:1
Total	46.38.90	Partial	463800	m:1
Total	46.39.10	Partial	463900	m:1
Total	46.39.20	Partial	463900	m:1
Total	46.41.10	Partial	464100	m:1
Total	46.41.20	Partial	464100	m:1
Total	46.41.90	Partial	464100	m:1
Total	46.42.10	Partial	464201	m:1
Total	46.42.20	Partial	464201	m:1
Total	46.42.30	Partial	464201	m:1
Total	46.42.40	Total	464202	1:1
Partial	46.43.10	Total	464301	m:n
Partial	46.43.10	Partial	464302	m:n
Total	46.43.20	Partial	464302	m:n
Total	46.43.30	Total	464303	1:1
Total	46.44.10	Partial	464400	m:1
Total	46.44.20	Partial	464400	m:1
Total	46.44.30	Partial	464400	m:1
Total	46.44.40	Partial	464400	m:1
Total	46.45.00	Total	464500	1:1
Total	46.46.10	Partial	464601	m:1
Total	46.46.20	Partial	464601	m:1
Total	46.46.30	Total	464602	1:1
Total	46.47.10	Partial	464700	m:1
Total	46.47.20	Partial	464700	m:1
Total	46.47.30	Partial	464700	m:1
Partial	46.48.00	Total	464801	1:n
Partial	46.48.00	Total	464802	1:n
Total	46.49.10	Partial	464901	m:1
Total	46.49.20	Partial	464901	m:1
Total	46.49.30	Total	464902	1:1
Total	46.49.40	Total	464903	1:1
Total	46.49.50	Total	464904	1:1
Partial	46.49.90	Total	464905	1:n

Coverage ATECO	ATECO 2022	Coverage NOGA	NOGA 2008	Link type
Partial	46.49.90	Total	464906	1:n
Partial	46.51.00	Total	465101	1:n
Partial	46.51.00	Total	465102	1:n
Total	46.52.01	Partial	465200	m:1
Total	46.52.02	Partial	465200	m:1
Total	46.52.09	Partial	465200	m:1
Total	46.61.00	Total	466100	1:1
Total	46.62.00	Total	466200	1:1
Total	46.63.00	Total	466300	1:1
Total	46.64.00	Total	466400	1:1
Total	46.65.00	Total	466500	1:1
Total	46.66.00	Total	466600	1:1
Total	46.69.11	Partial	466900	m:1
Total	46.69.19	Partial	466900	m:1
Total	46.69.20	Partial	466900	m:1
Total	46.69.30	Partial	466900	m:1
Total	46.69.91	Partial	466900	m:1
Total	46.69.92	Partial	466900	m:1
Total	46.69.93	Partial	466900	m:1
Total	46.69.94	Partial	466900	m:1
Total	46.69.99	Partial	466900	m:1
Total	46.71.00	Total	467100	1:1
Total	46.72.10	Partial	467200	m:1
Total	46.72.20	Partial	467200	m:1
Total	46.73.10	Total	467301	1:1
Total	46.73.21	Partial	467302	m:1
Total	46.73.22	Partial	467302	m:1
Total	46.73.23	Partial	467302	m:1
Total	46.73.29	Partial	467302	m:1
Total	46.73.30	Total	467303	1:1
Total	46.73.40	Partial	467302	m:1
Total	46.74.10	Partial	467400	m:1
Total	46.74.20	Partial	467400	m:1
Total	46.75.01	Partial	467500	m:1
Total	46.75.02	Partial	467500	m:1
Total	46.76.10	Partial	467600	m:1
Total	46.76.20	Partial	467600	m:1
Total	46.76.30	Partial	467600	m:1
Total	46.76.90	Partial	467600	m:1
Partial	46.77.10	Total	467701	m:n
Partial	46.77.10	Partial	467702	m:n
Total	46.77.20	Partial	467702	m:n
Total	46.90.00	Total	469000	1:1
Total	47.11.10	Partial	471101	m:n

Coverage ATECO	ATECO 2022	Coverage NOGA	NOGA 2008	Link type
Partial	47.11.20	Partial	471102	m:n
Partial	47.11.20	Partial	471103	m:n
Partial	47.11.30	Partial	471101	m:n
Partial	47.11.30	Partial	471102	m:n
Partial	47.11.30	Partial	471103	m:n
Partial	47.11.30	Partial	471104	m:n
Partial	47.11.30	Partial	471105	m:n
Partial	47.11.40	Partial	471104	m:n
Partial	47.11.40	Partial	471105	m:n
Partial	47.11.50	Partial	471101	m:n
Partial	47.11.50	Partial	471102	m:n
Partial	47.11.50	Partial	471103	m:n
Partial	47.11.50	Partial	471104	m:n
Partial	47.11.50	Partial	471105	m:n
Partial	47.19.10	Total	471901	m:n
Partial	47.19.10	Partial	471902	m:n
Total	47.19.20	Partial	471902	m:n
Total	47.19.90	Partial	471902	m:n
Total	47.21.01	Partial	472100	m:1
Total	47.21.02	Partial	472100	m:1
Total	47.22.00	Total	472200	1:1
Total	47.23.00	Total	472300	1:1
Partial	47.24.10	Partial	472401	m:n
Partial	47.24.10	Partial	472402	m:n
Partial	47.24.20	Partial	472401	m:n
Partial	47.24.20	Partial	472402	m:n
Total	47.25.00	Total	472500	1:1
Total	47.26.00	Total	472600	1:1
Total	47.29.10	Total	472901	1:1
Total	47.29.20	Partial	472902	m:1
Total	47.29.30	Partial	472902	m:1
Total	47.29.90	Partial	472902	m:1
Total	47.30.00	Total	473000	1:1
Total	47.41.00	Total	474100	1:1
Total	47.42.00	Total	474200	1:1
Total	47.43.00	Total	474300	1:1
Total	47.51.10	Partial	475100	m:1
Total	47.51.20	Partial	475100	m:1
Total	47.52.10	Total	475201	1:1
Total	47.52.20	Partial	475202	m:1
Total	47.52.30	Partial	475202	m:1
Total	47.52.40	Partial	475202	m:1
Total	47.53.11	Partial	475300	m:1
Total	47.53.12	Partial	475300	m:1

Coverage ATECO	ATECO 2022	Coverage NOGA	NOGA 2008	Link type
Total	47.53.20	Partial	475300	m:1
Total	47.54.00	Total	475400	1:1
Total	47.59.10	Total	475902	1:1
Total	47.59.20	Partial	475903	m:1
Total	47.59.30	Partial	475903	m:1
Total	47.59.40	Partial	475903	m:1
Total	47.59.50	Partial	475903	m:1
Total	47.59.60	Total	475901	1:1
Total	47.59.91	Partial	475903	m:1
Total	47.59.99	Partial	475903	m:1
Total	47.61.00	Total	476100	1:1
Total	47.62.10	Total	476201	1:1
Total	47.62.20	Total	476202	1:1
Total	47.63.00	Total	476300	1:1
Partial	47.64.10	Total	476401	m:n
Partial	47.64.10	Partial	476402	m:n
Total	47.64.20	Partial	476402	m:n
Total	47.65.00	Total	476500	1:1
Partial	47.71.10	Partial	477101	m:n
Partial	47.71.10	Partial	477102	m:n
Partial	47.71.10	Partial	477103	m:n
Partial	47.71.20	Partial	477101	m:n
Partial	47.71.20	Partial	477102	m:n
Partial	47.71.20	Partial	477103	m:n
Partial	47.71.30	Partial	477101	m:n
Partial	47.71.30	Partial	477102	m:n
Partial	47.71.30	Partial	477103	m:n
Partial	47.71.40	Partial	477101	m:n
Partial	47.71.40	Partial	477102	m:n
Partial	47.71.40	Partial	477103	m:n
Partial	47.71.40	Partial	477104	m:n
Total	47.71.50	Total	477105	1:1
Total	47.72.10	Total	477201	1:1
Total	47.72.20	Total	477202	1:1
Total	47.73.10	Partial	477300	m:1
Total	47.73.20	Partial	477300	m:1
Total	47.74.00	Total	477400	1:1
Total	47.75.10	Total	477502	1:1
Total	47.75.20	Total	477501	1:1
Partial	47.76.10	Total	477601	1:n
Partial	47.76.10	Total	477602	1:n
Total	47.76.20	Total	477603	1:1
Total	47.77.00	Total	477700	1:1
Total	47.78.10	Partial	477806	m:n

Coverage ATECO	ATECO 2022	Coverage NOGA	NOGA 2008	Link type
Partial	47.78.20	Total	477802	1:n
Partial	47.78.20	Total	477803	1:n
Total	47.78.31	Total	477805	1:1
Total	47.78.32	Partial	477804	m:n
Total	47.78.33	Partial	477804	m:n
Total	47.78.34	Partial	477804	m:n
Total	47.78.35	Partial	477806	m:n
Partial	47.78.36	Partial	477804	m:n
Partial	47.78.36	Partial	477806	m:n
Total	47.78.37	Partial	477806	m:n
Total	47.78.40	Total	477801	1:1
Total	47.78.50	Partial	477806	m:n
Total	47.78.60	Partial	477806	m:n
Total	47.78.91	Partial	477806	m:n
Total	47.78.92	Partial	477806	m:n
Total	47.78.93	Partial	477806	m:n
Total	47.78.94	Partial	477806	m:n
Total	47.78.99	Partial	477806	m:n
Total	47.79.10	Partial	477902	m:1
Total	47.79.20	Total	477901	1:1
Total	47.79.30	Partial	477902	m:1
Total	47.79.40	Partial	477902	m:1
Total	47.81.01	Partial	478100	m:1
Total	47.81.02	Partial	478100	m:1
Total	47.81.03	Partial	478100	m:1
Total	47.81.09	Partial	478100	m:1
Total	47.82.01	Partial	478200	m:1
Total	47.82.02	Partial	478200	m:1
Total	47.89.01	Partial	478900	m:1
Total	47.89.02	Partial	478900	m:1
Total	47.89.03	Partial	478900	m:1
Total	47.89.04	Partial	478900	m:1
Total	47.89.05	Partial	478900	m:1
Total	47.89.09	Partial	478900	m:1
Total	47.91.10	Partial	479100	m:1
Total	47.91.20	Partial	479100	m:1
Total	47.91.30	Partial	479100	m:1
Total	47.99.10	Partial	479900	m:1
Total	47.99.20	Partial	479900	m:1
Total	49.10.00	Total	491000	1:1
Total	49.20.00	Total	492000	1:1
Total	49.31.00	Total	493100	1:1
Total	49.32.10	Partial	493200	m:1
Total	49.32.20	Partial	493200	m:1

Coverage ATECO	ATECO 2022	Coverage NOGA	NOGA 2008	Link type
Total	49.39.01	Total	493903	1:1
Partial	49.39.09	Total	493901	1:n
Partial	49.39.09	Total	493902	1:n
Total	49.41.00	Total	494100	1:1
Total	49.42.00	Total	494200	1:1
Total	49.50.10	Partial	495000	m:1
Total	49.50.20	Partial	495000	m:1
Total	50.10.00	Total	501000	1:1
Total	50.20.00	Total	502000	1:1
Total	50.30.00	Total	503000	1:1
Total	50.40.00	Total	504000	1:1
Total	51.10.10	Partial	511000	m:1
Total	51.10.20	Partial	511000	m:1
Total	51.21.00	Total	512100	1:1
Total	51.22.00	Total	512200	1:1
Total	52.10.10	Partial	521000	m:1
Total	52.10.20	Partial	521000	m:1
Total	52.21.10	Partial	522100	m:1
Total	52.21.20	Partial	522100	m:1
Total	52.21.30	Partial	522100	m:1
Total	52.21.40	Partial	522100	m:1
Total	52.21.50	Partial	522100	m:1
Total	52.21.60	Partial	522100	m:1
Total	52.21.90	Partial	522100	m:1
Total	52.22.01	Partial	522200	m:1
Total	52.22.09	Partial	522200	m:1
Total	52.23.00	Total	522300	1:1
Total	52.24.10	Partial	522400	m:1
Total	52.24.20	Partial	522400	m:1
Total	52.24.30	Partial	522400	m:1
Total	52.24.40	Partial	522400	m:1
Total	52.29.10	Partial	522900	m:1
Total	52.29.21	Partial	522900	m:1
Total	52.29.22	Partial	522900	m:1
Total	53.10.00	Total	531000	1:1
Total	53.20.00	Total	532000	1:1
Partial	55.10.00	Total	551001	1:n
Partial	55.10.00	Total	551002	1:n
Partial	55.10.00	Total	551003	1:n
Partial	55.20.10	Partial	552002	m:n
Partial	55.20.10	Partial	552003	m:n
Partial	55.20.20	Partial	552002	m:n
Partial	55.20.20	Partial	552003	m:n
Partial	55.20.30	Partial	552002	m:n

Coverage ATECO	ATECO 2022	Coverage NOGA	NOGA 2008	Link type
Partial	55.20.30	Partial	552003	m:n
Partial	55.20.40	Partial	552002	m:n
Partial	55.20.40	Partial	552003	m:n
Partial	55.20.51	Total	552001	m:n
Partial	55.20.51	Partial	552003	m:n
Partial	55.20.52	Partial	552002	m:n
Partial	55.20.52	Partial	552003	m:n
Partial	55.20.53	Partial	552002	m:n
Partial	55.20.53	Partial	552003	m:n
Partial	55.30.00	Total	553001	1:n
Partial	55.30.00	Total	553002	1:n
Total	55.90.10	Partial	559000	m:1
Total	55.90.20	Partial	559000	m:1
Partial	56.10.11	Partial	561001	m:n
Partial	56.10.11	Partial	561002	m:n
Partial	56.10.11	Partial	561003	m:n
Partial	56.10.12	Partial	561001	m:n
Partial	56.10.12	Partial	561002	m:n
Partial	56.10.12	Partial	561003	m:n
Partial	56.10.13	Partial	561001	m:n
Partial	56.10.13	Partial	561002	m:n
Partial	56.10.13	Partial	561003	m:n
Partial	56.10.20	Partial	561001	m:n
Partial	56.10.20	Partial	561003	m:n
Partial	56.10.30	Partial	561001	m:n
Partial	56.10.30	Partial	561003	m:n
Partial	56.10.41	Partial	561001	m:n
Partial	56.10.41	Partial	561003	m:n
Partial	56.10.42	Partial	561001	m:n
Partial	56.10.42	Partial	561003	m:n
Partial	56.10.50	Partial	561001	m:n
Partial	56.10.50	Partial	561003	m:n
Total	56.21.00	Total	562100	1:1
Total	56.29.10	Partial	562900	m:1
Total	56.29.20	Partial	562900	m:1
Partial	56.30.00	Total	563001	1:n
Partial	56.30.00	Total	563002	1:n
Total	58.11.00	Total	581100	1:1
Total	58.12.01	Partial	581200	m:1
Total	58.12.02	Partial	581200	m:1
Total	58.13.00	Total	581300	1:1
Total	58.14.00	Total	581400	1:1
Total	58.19.00	Total	581900	1:1
Total	58.21.00	Total	582100	1:1

Coverage ATECO	ATECO 2022	Coverage NOGA	NOGA 2008	Link type
Total	58.29.00	Total	582900	1:1
Total	59.11.00	Total	591100	1:1
Total	59.12.00	Total	591200	1:1
Total	59.13.00	Total	591300	1:1
Total	59.14.00	Total	591400	1:1
Total	59.20.10	Partial	592000	m:1
Total	59.20.20	Partial	592000	m:1
Total	59.20.30	Partial	592000	m:1
Total	60.10.00	Total	601000	1:1
Total	60.20.00	Total	602000	1:1
Total	61.10.00	Total	611000	1:1
Total	61.20.00	Total	612000	1:1
Total	61.30.00	Total	613000	1:1
Total	61.90.10	Partial	619000	m:1
Total	61.90.20	Partial	619000	m:1
Total	61.90.91	Partial	619000	m:1
Total	61.90.99	Partial	619000	m:1
Total	62.01.00	Total	620100	1:1
Total	62.02.00	Total	620200	1:1
Total	62.03.00	Total	620300	1:1
Total	62.09.01	Partial	620900	m:1
Total	62.09.09	Partial	620900	m:1
Total	63.11.11	Partial	631100	m:1
Total	63.11.19	Partial	631100	m:1
Total	63.11.20	Partial	631100	m:1
Total	63.11.30	Partial	631100	m:1
Total	63.12.00	Total	631200	1:1
Total	63.91.00	Total	639100	1:1
Total	63.99.00	Total	639900	1:1
Total	64.11.00	Total	641100	1:1
Partial	64.19.10	Total	641901	1:n
Partial	64.19.10	Total	641902	1:n
Partial	64.19.10	Total	641903	1:n
Partial	64.19.10	Total	641904	1:n
Partial	64.19.10	Total	641905	1:n
Partial	64.19.10	Total	641906	1:n
Partial	64.19.10	Total	641907	1:n
Partial	64.19.10	Total	641908	1:n
Partial	64.19.10	Total	641909	1:n
Partial	64.19.10	Total	641910	1:n
Partial	64.19.10	Total	641911	1:n
Total	64.19.20	Partial	641912	m:1
Total	64.19.30	Partial	641912	m:1
Total	64.19.40	Partial	641912	m:1

Coverage ATECO	**ATECO 2022**	**Coverage NOGA**	**NOGA 2008**	**Link type**
Partial	64.20.00	Total	642001	1:n
Partial	64.20.00	Total	642002	1:n
Total	64.30.10	Partial	643000	m:1
Total	64.30.20	Partial	643000	m:1
Total	64.91.00	Total	649100	1:1
Total	64.92.01	Partial	649202	m:n
Partial	64.92.09	Total	649201	m:n
Partial	64.92.09	Partial	649202	m:n
Total	64.99.10	Total	649901	1:1
Total	64.99.20	Partial	649903	m:n
Total	64.99.30	Partial	649903	m:n
Total	64.99.40	Partial	649903	m:n
Total	64.99.50	Partial	649903	m:n
Partial	64.99.60	Total	649902	m:n
Partial	64.99.60	Partial	649903	m:n
Total	65.11.00	Total	651100	1:1
Partial	65.12.00	Total	651201	1:n
Partial	65.12.00	Total	651202	1:n
Partial	65.12.00	Total	651203	1:n
Partial	65.12.00	Total	651204	1:n
Total	65.20.00	Total	652000	1:1
Total	65.30.10	Partial	653000	m:1
Total	65.30.20	Partial	653000	m:1
Total	65.30.30	Partial	653000	m:1
Total	66.11.00	Total	661100	1:1
Total	66.12.00	Total	661200	1:1
Total	66.19.10	Partial	661900	m:1
Total	66.19.21	Partial	661900	m:1
Total	66.19.22	Partial	661900	m:1
Total	66.19.30	Partial	661900	m:1
Total	66.19.40	Partial	641912	m:1
Total	66.19.50	Partial	661900	m:1
Total	66.21.00	Total	662100	1:1
Total	66.22.01	Partial	662200	m:1
Total	66.22.02	Partial	662200	m:1
Total	66.22.03	Partial	662200	m:1
Total	66.22.04	Partial	662200	m:1
Partial	66.29.01	Total	662901	m:n
Partial	66.29.01	Partial	662902	m:n
Total	66.29.09	Partial	662902	m:n
Partial	66.30.00	Total	663001	1:n
Partial	66.30.00	Total	663002	1:n
Total	68.10.00	Total	681000	1:1
Partial	68.20.01	Total	682001	m:n

Coverage ATECO	**ATECO 2022**	**Coverage NOGA**	**NOGA 2008**	**Link type**
Partial	68.20.01	Partial	682002	m:n
Total	68.20.02	Partial	682002	m:n
Total	68.31.00	Total	683100	1:1
Total	68.32.00	Total	683200	1:1
Partial	69.10.10	Partial	691001	m:n
Partial	69.10.10	Partial	691002	m:n
Partial	69.10.20	Partial	691001	m:n
Partial	69.10.20	Partial	691002	m:n
Total	69.20.11	Partial	692000	m:1
Total	69.20.12	Partial	692000	m:1
Total	69.20.13	Partial	692000	m:1
Total	69.20.14	Partial	692000	m:1
Total	69.20.15	Partial	692000	m:1
Total	69.20.20	Partial	692000	m:1
Total	69.20.30	Partial	692000	m:1
Partial	70.10.00	Total	701001	1:n
Partial	70.10.00	Total	701002	1:n
Total	70.21.00	Total	702100	1:1
Total	70.22.01	Partial	702200	m:1
Total	70.22.09	Partial	702200	m:1
Partial	71.11.00	Total	711101	1:n
Partial	71.11.00	Total	711102	1:n
Partial	71.11.00	Total	711103	1:n
Partial	71.12.10	Total	711201	m:n
Partial	71.12.10	Total	711202	m:n
Partial	71.12.10	Partial	711205	m:n
Total	71.12.20	Total	711203	1:1
Total	71.12.30	Total	711204	1:1
Total	71.12.40	Partial	711205	m:n
Total	71.12.50	Partial	711205	m:n
Total	71.20.10	Partial	712000	m:1
Total	71.20.21	Partial	712000	m:1
Total	71.20.22	Partial	712000	m:1
Total	71.20.23	Partial	712000	m:1
Total	72.11.00	Total	721100	1:1
Total	72.19.01	Partial	721900	m:1
Total	72.19.09	Partial	721900	m:1
Total	72.20.00	Total	722000	1:1
Total	73.11.01	Partial	731100	m:1
Total	73.11.02	Partial	731100	m:1
Total	73.12.00	Total	731200	1:1
Total	73.20.00	Total	732000	1:1
Total	74.10.10	Total	741001	1:1
Total	74.10.21	Partial	741002	m:1

Coverage ATECO	ATECO 2022	Coverage NOGA	NOGA 2008	Link type
Total	74.10.29	Partial	741002	m:1
Total	74.10.30	Partial	741003	m:1
Total	74.10.90	Partial	741003	m:1
Total	74.20.11	Partial	742001	m:1
Total	74.20.12	Partial	742001	m:1
Total	74.20.19	Partial	742001	m:1
Total	74.20.20	Total	742002	1:1
Total	74.30.00	Total	743000	1:1
Total	74.90.11	Partial	749000	m:1
Total	74.90.12	Partial	749000	m:1
Total	74.90.13	Partial	749000	m:1
Total	74.90.14	Partial	749000	m:1
Total	74.90.21	Partial	749000	m:1
Total	74.90.29	Partial	749000	m:1
Total	74.90.31	Partial	749000	m:1
Total	74.90.32	Partial	749000	m:1
Total	74.90.33	Partial	749000	m:1
Total	74.90.91	Partial	749000	m:1
Total	74.90.92	Partial	749000	m:1
Total	74.90.93	Partial	749000	m:1
Total	74.90.94	Partial	749000	m:1
Total	74.90.99	Partial	749000	m:1
Total	75.00.00	Total	750000	1:1
Total	77.11.00	Total	771100	1:1
Total	77.12.00	Total	771200	1:1
Total	77.21.01	Partial	772100	m:1
Total	77.21.02	Partial	772100	m:1
Total	77.21.09	Partial	772100	m:1
Total	77.22.00	Total	772200	1:1
Total	77.29.10	Partial	772900	m:1
Total	77.29.90	Partial	772900	m:1
Total	77.31.00	Total	773100	1:1
Total	77.32.00	Total	773200	1:1
Total	77.33.00	Total	773300	1:1
Total	77.34.00	Total	773400	1:1
Total	77.35.00	Total	773500	1:1
Total	77.39.10	Partial	773900	m:1
Total	77.39.91	Partial	773900	m:1
Total	77.39.92	Partial	773900	m:1
Total	77.39.93	Partial	773900	m:1
Total	77.39.94	Partial	773900	m:1
Total	77.39.99	Partial	773900	m:1
Total	77.40.00	Total	774000	1:1
Total	78.10.00	Total	781000	1:1

Coverage ATECO	ATECO 2022	Coverage NOGA	NOGA 2008	Link type
Total	78.20.00	Total	782000	1:1
Total	78.30.00	Total	783000	1:1
Total	79.11.00	Total	791100	1:1
Total	79.12.00	Total	791200	1:1
Total	79.90.11	Partial	799002	m:n
Partial	79.90.19	Total	799001	m:n
Partial	79.90.19	Partial	799002	m:n
Total	79.90.20	Partial	799002	m:n
Total	80.10.00	Total	801000	1:1
Total	80.20.00	Total	802000	1:1
Total	80.30.00	Total	803000	1:1
Total	81.10.00	Total	811000	1:1
Total	81.21.00	Total	812100	1:1
Total	81.22.01	Partial	812202	m:n
Partial	81.22.02	Total	812201	m:n
Partial	81.22.02	Partial	812202	m:n
Total	81.29.10	Partial	812900	m:1
Total	81.29.91	Partial	812900	m:1
Total	81.29.99	Partial	812900	m:1
Total	81.30.00	Total	813000	1:1
Total	82.11.01	Partial	821100	m:1
Total	82.11.02	Partial	821100	m:1
Total	82.19.01	Total	821901	1:1
Total	82.19.09	Total	821902	1:1
Total	82.20.00	Total	822000	1:1
Total	82.30.00	Total	823000	1:1
Total	82.91.10	Partial	829100	m:1
Total	82.91.20	Partial	829100	m:1
Total	82.92.10	Partial	829200	m:1
Total	82.92.20	Partial	829200	m:1
Total	82.99.10	Partial	829900	m:1
Total	82.99.20	Partial	829900	m:1
Total	82.99.30	Partial	829900	m:1
Total	82.99.40	Partial	829900	m:1
Total	82.99.91	Partial	829900	m:1
Total	82.99.99	Partial	829900	m:1
Total	84.11.10	Partial	841100	m:1
Total	84.11.20	Partial	841100	m:1
Total	84.12.10	Partial	841200	m:1
Total	84.12.20	Partial	841200	m:1
Total	84.12.30	Partial	841200	m:1
Total	84.12.40	Partial	841200	m:1
Total	84.13.10	Partial	841300	m:1
Total	84.13.20	Partial	841300	m:1

Coverage ATECO	ATECO 2022	Coverage NOGA	NOGA 2008	Link type
Total	84.13.30	Partial	841300	m:1
Total	84.13.40	Partial	841300	m:1
Total	84.13.50	Partial	841300	m:1
Total	84.13.60	Partial	841300	m:1
Total	84.13.70	Partial	841300	m:1
Total	84.13.80	Partial	841300	m:1
Total	84.13.90	Partial	841300	m:1
Total	84.21.00	Total	842100	1:1
Total	84.22.00	Total	842201	1:1
Partial	84.23.00	Total	842301	1:n
Partial	84.23.00	Total	842302	1:n
Total	84.24.00	Total	842400	1:1
Total	84.25.10	Total	842500	1:1
Total	84.25.20	Total	842202	1:1
Total	84.30.00	Total	843000	1:1
Total	85.10.00	Total	851000	1:1
Partial	85.20.00	Total	852001	1:n
Partial	85.20.00	Total	852002	1:n
Partial	85.20.00	Total	852003	1:n
Total	85.31.10	Total	853101	1:1
Partial	85.31.20	Total	853102	1:n
Partial	85.31.20	Total	853103	1:n
Total	85.32.01	Partial	853200	m:1
Total	85.32.02	Partial	853200	m:1
Total	85.32.03	Partial	853200	m:1
Total	85.32.09	Partial	853200	m:1
Total	85.41.00	Total	854100	1:1
Partial	85.42.00	Total	854201	1:n
Partial	85.42.00	Total	854202	1:n
Partial	85.42.00	Total	854203	1:n
Total	85.51.00	Total	855100	1:1
Total	85.52.01	Partial	855200	m:1
Total	85.52.09	Partial	855200	m:1
Total	85.53.00	Total	855300	1:1
Total	85.59.10	Partial	855903	m:1
Total	85.59.20	Partial	855903	m:1
Total	85.59.30	Total	855901	1:1
Partial	85.59.90	Total	855902	1:n
Partial	85.59.90	Total	855904	1:n
Total	85.60.01	Partial	856000	m:1
Total	85.60.09	Partial	856000	m:1
Total	86.10.10	Partial	861001	m:1
Total	86.10.20	Partial	861002	m:1
Total	86.10.30	Partial	861001	m:1

Coverage ATECO	ATECO 2022	Coverage NOGA	NOGA 2008	Link type
Total	86.10.40	Partial	861002	m:1
Total	86.21.00	Total	862100	1:1
Total	86.22.01	Partial	862200	m:1
Total	86.22.02	Partial	862200	m:1
Total	86.22.03	Partial	862200	m:1
Total	86.22.04	Partial	862200	m:1
Total	86.22.05	Partial	862200	m:1
Total	86.22.06	Partial	862200	m:1
Total	86.22.09	Partial	862200	m:1
Total	86.23.00	Total	862300	1:1
Total	86.90.11	Partial	869006	m:1
Total	86.90.12	Partial	869006	m:1
Total	86.90.13	Partial	869006	m:1
Total	86.90.21	Total	869002	1:1
Partial	86.90.29	Total	869003	1:n
Partial	86.90.29	Total	869004	1:n
Partial	86.90.29	Total	869005	1:n
Total	86.90.30	Total	869001	1:1
Total	86.90.41	Partial	869007	m:1
Total	86.90.42	Partial	869007	m:1
Total	87.10.00	Total	871000	1:1
Partial	87.20.00	Total	872001	1:n
Partial	87.20.00	Total	872002	1:n
Partial	87.30.00	Total	873001	1:n
Partial	87.30.00	Total	873002	1:n
Partial	87.90.00	Total	879001	1:n
Partial	87.90.00	Total	879002	1:n
Partial	87.90.00	Total	879003	1:n
Total	88.10.00	Total	881000	1:1
Total	88.91.00	Total	889100	1:1
Partial	88.99.00	Total	889901	1:n
Partial	88.99.00	Total	889902	1:n
Total	90.01.01	Partial	900101	m:n
Partial	90.01.09	Partial	900101	m:n
Partial	90.01.09	Total	900102	m:n
Total	90.02.01	Partial	900200	m:1
Total	90.02.02	Partial	900200	m:1
Total	90.02.09	Partial	900200	m:1
Total	90.03.01	Total	900303	1:1
Partial	90.03.02	Total	900301	m:n
Partial	90.03.02	Partial	900302	m:n
Total	90.03.09	Partial	900302	m:n
Total	90.04.00	Total	900400	1:1
Total	91.01.00	Total	910100	1:1

Coverage ATECO	ATECO 2022	Coverage NOGA	NOGA 2008	Link type
Total	91.02.00	Total	910200	1:1
Total	91.03.00	Total	910300	1:1
Total	91.04.00	Total	910400	1:1
Total	92.00.01	Partial	920000	m:1
Total	92.00.02	Partial	920000	m:1
Total	92.00.09	Partial	920000	m:1
Total	93.11.10	Partial	931100	m:1
Total	93.11.20	Partial	931100	m:1
Total	93.11.30	Partial	931100	m:1
Total	93.11.90	Partial	931100	m:1
Total	93.12.00	Total	931200	1:1
Total	93.13.00	Total	931300	1:1
Total	93.19.10	Partial	931900	m:1
Total	93.19.91	Partial	931900	m:1
Total	93.19.92	Partial	931900	m:1
Total	93.19.99	Partial	931900	m:1
Total	93.21.01	Partial	932100	m:1
Total	93.21.02	Partial	932100	m:1
Total	93.29.10	Partial	932900	m:1
Total	93.29.20	Partial	932900	m:1
Total	93.29.30	Partial	932900	m:1
Total	93.29.90	Partial	932900	m:1
Total	94.11.00	Total	941100	1:1
Total	94.12.10	Partial	941200	m:1
Total	94.12.20	Partial	941200	m:1
Total	94.20.00	Total	942000	1:1
Partial	94.91.00	Total	949101	1:n
Partial	94.91.00	Total	949102	1:n
Total	94.92.00	Total	949200	1:1
Total	94.99.10	Partial	949901	m:1
Total	94.99.20	Partial	949901	m:1
Total	94.99.30	Partial	949904	m:n
Total	94.99.40	Partial	949904	m:n
Total	94.99.50	Partial	949901	m:1
Total	94.99.60	Partial	949904	m:n
Partial	94.99.90	Total	949902	m:n
Partial	94.99.90	Total	949903	m:n
Partial	94.99.90	Partial	949904	m:n
Total	95.11.00	Total	951100	1:1
Total	95.12.01	Partial	951200	m:1
Total	95.12.09	Partial	951200	m:1
Total	95.21.00	Total	952100	1:1
Total	95.22.01	Partial	952200	m:1
Total	95.22.02	Partial	952200	m:1

Coverage ATECO	**ATECO 2022**	**Coverage NOGA**	**NOGA 2008**	**Link type**
Total	95.23.00	Total	952300	1:1
Total	95.24.01	Partial	952400	m:1
Total	95.24.02	Partial	952400	m:1
Total	95.25.00	Total	952500	1:1
Total	95.29.01	Partial	952900	m:1
Total	95.29.02	Partial	952900	m:1
Total	95.29.03	Partial	952900	m:1
Total	95.29.04	Partial	952900	m:1
Total	95.29.09	Partial	952900	m:1
Partial	96.01.10	Partial	960101	m:n
Partial	96.01.10	Partial	960102	m:n
Partial	96.01.20	Partial	960101	m:n
Partial	96.01.20	Partial	960102	m:n
Partial	96.01.30	Partial	960101	m:n
Partial	96.01.30	Partial	960102	m:n
Total	96.02.01	Partial	960201	m:n
Total	96.02.02	Partial	960202	m:n
Total	96.02.03	Partial	960202	m:n
Total	96.03.00	Total	960300	1:1
Partial	96.04.10	Total	960401	m:n
Partial	96.04.10	Partial	960402	m:n
Total	96.04.20	Partial	960402	m:n
Total	96.09.01	Partial	960900	m:1
Total	96.09.02	Partial	960900	m:1
Total	96.09.03	Partial	960900	m:1
Total	96.09.04	Partial	960900	m:1
Total	96.09.05	Partial	960900	m:1
Total	96.09.09	Partial	960900	m:1
Total	97.00.01	Partial	970000	m:1
Total	97.00.02	Partial	970000	m:1
Total	98.10.00	Total	981000	1:1
Total	98.20.00	Total	982000	1:1
Partial	99.00.00	Total	990001	1:n
Partial	99.00.00	Total	990002	1:n
Partial	99.00.00	Total	990003	1:n

Source: Author's processing

References

Alonzi, F., S. Ambroselli, and A. Valery. 2020*a*. "COVID-19 emergency: a comprehensive overview of the economic classification systems for the manufacturing industry". *Rivista di statistica ufficiale/Review of official statistics*, N. 3/2020: 51-117. https://www.doi.org/10.1481/ISTATRIVISTASTATISTICAUFFICIALE_3.2020.02.

Alonzi, F., S. Ambroselli, and A. Valery. 2020*b*. "COVID-19 emergency and the impact on classification systems: a preliminary focus on manufacturing industry". *Istat working papers*, N. 7/2020. Roma, Italy: Istat. https://www.istat.it/en/publication/covid-19-emergency-and-the-impact-on-classification-systems-a-preliminary-focus-on-manufacturing-industry-2/.

Alonzi, F., and C. Viviano. 2022. "The new updating process of the ATECO classification". Paper prepared for the IAOS *Conference, Kraków*, 26-28 April 2022. https://www.iaos2022.pl/.

Eurostat. 2021. "European Business Statistics Manual 2021 edition". *Manuals and guidelines*. Luxembourg: Publications Office of the European Union. https://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/ks-gq-21-001.

Eurostat. 2019. *Formalization of the Structure and Content of Statistical Classifications*. Version 1.0, 1st October 2019. Luxembourg: Eurostat.

Eurostat. 2008. "NACE Rev.2. Statistical classification of economic activities in the European Community". *Methodologies and Working papers*. Luxembourg: Office for Official Publications of the European Communities. https://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/ks-ra-07-015.

Hancock, A. 2013. "Best Practice Guidelines for Developing International Statistical Classifications". Paper prepared for the *Meeting of the Expert Group on International Statistical Classifications*, 13th - 15th May 2013, New York, NY, U.S.

Hoffmann, E., and M. Chamie. 1999. "Standard Statistical Classifications: Basic Principles". Paper prepared for the *30th Session of the Statistical Commission*, 1st - 5th March 1999, New York, NY, U.S.

Istituto Nazionale di Statistica - Istat. 2022. *Classificazione delle attività economiche - Ateco 2007*. Roma, Italia: Istat. https://www.istat.it/classificazione/classificazione-delle-attivita-economiche-ateco/.

Istituto Nazionale di Statistica - Istat. 2009. "Classificazione delle attività economiche Ateco 2007 - derivata dalla Nace Rev. 2". *Metodi e Norme*. Roma, Italia: Istat. https://www4.istat.it/it/files/2011/03/metenorme09_40classificazione_attivita_economiche_2007.pdf.

Swiss Federal Statistical Office. 2008. NOGA 2008. *General Classification of Economic Activities Explanatory notes*. Neuchâtel, Switzerland: Federal Statistical Office. https://www.bfs.admin.ch/bfs/en/home/statistics/industry-services/nomenclatures/noga.html.

United Nations Statistical Division - UNSD, Department of Economic and Social Affairs. 2008. "International Standard Industrial Classification of All Economic Activities. Revision 4". *Statistical Papers*, Series M, N. 4/Rev. 4. New York, NY, U.S.: United Nations Publications. https://unstats.un.org/unsd/publication/seriesm/seriesm_4rev4e.pdf.

Business confidence indicators across (similar) surveys

Tatiana Cesaroni[1], Libero Monteforte[2]

Abstract

Business survey indicators are widely used to monitor and forecast in the short term the business outlook. The main surveys are collected by different organisations, such as national statistical offices, private research entities and central banks. In Italy, the Italian National Institute of Statistics (Istat) collects monthly survey indicators, as part of the European Commission survey programme and business survey data are also gathered by the Bank of Italy, using a distinct sample.
This study evaluates the similarities and differences between two similar business survey indicators, collected by Istat and the Bank of Italy, on employment expectations and the general economic conditions. We compare the cyclical properties and ability of the two indicators to forecast key macroeconomic variables such as industrial production and employment. In addition to the standard indicators (balances between positive and negative judgements), we explore the usefulness of the single-tail components (i.e. positive and negative judgements) to track the business cycle. We find that the series of the two surveys has a significant and similar predictive power, and, with a few exceptions, it is not possible to exclude one of the indicators completely; the optimal strategy for nowcasting is therefore to use both, as in a forecast combination.

Keywords: Business survey data, forecasting accuracy, VAR models, nowcasting, macroeconomic uncertainty.

JEL Classification: C32, E32.

DOI: https://doi.org/10.1481/ISTATRIVISTASTATISTICAUFFICIALE_1-2-3.2023.05.

1 Tatiana Cesaroni (tatiana.cesaroni@mef.gov.it), Italian Ministry of Economy and Finance - MEF.
2 Libero Monteforte (libero.monteforte@upbilancio.it), Italian Parliamentary Budget Office.

The views and opinions expressed are those of the authors and do not necessarily reflect the official policy or position of the Italian National Institute of Statistics - Istat.
The authors would like to thank the anonymous reviewers for their comments and suggestions, which enhanced the quality of this article.

1. Introduction[3]

Business confidence indicators, which are widely used by policymakers to track economic outlook and forecasting, are collected by different organisations such as national statistical offices, private research entities and central banks. In Italy, the official business confidence indicators are collected by the Italian National Institute of Statistics (Istat), in the context of the European Commission (EC) business survey programme. Since 1999, some of these indicators have also been collected by the Bank of Italy (BoI), within the Survey on Inflation and Growth Expectations (SIGE).

This study aims to compare the most similar business confidence indicators gathered by Istat and the BoI, namely the sentiment indicator on the general economic situation and the expectations of employment in the next three months. These two indicators are produced through the aggregation of qualitative data through the so-called "balance," given by the difference between positive and negative answers provided by firms.

We checked all the business survey indicators available in the two surveys. We ended up focussing only on those two, which refer to the same economic phenomenon and have a similar quantitative assessment. For example, we do not consider the indicator regarding the price expectations, despite the similar wording of the question, because Istat collects qualitative information (in the form of a balance given by the difference between positive and negative judgements). At the same time, the BoI asks for the expected percentage change in the price.

The novelty of this study is the comparison of forecasting power between indicators capturing the same economic phenomenon from two similar business surveys. To the best of our knowledge, this is new in the literature. More in detail, after a discussion of the statistical characteristics and properties of the two indicators (Istat and BoI), we assess their forecast performance using multivariate models. We also explore the potential usefulness of positive and negative judgements, individually considered, to track the business cycle. Indeed, an increase in positive answers from respondents could be associated with an expansionary business cycle phase; quite the opposite, an increase in negative judgements of the firms is expected to be associated with a worsening in macroeconomic conditions; these

3 The authors would like to thank all the participants in the Bank of Italy Lunch Seminar held on 18 November 2017, and the participants in the CFE ERCIM Conference held on 15-17 December 2017 in London for their valuable suggestions.

tail components may provide a cleaner signal at certain phases of the business cycle, but by losing observations, the signal may also be more unstable; therefore, the question of its usefulness in forecasting is empirical. Finally, we consider the possibility of using these business surveys to track macroeconomic uncertainty, following the approach proposed by Bachmann *et al.* (2013).

Many studies have attempted to analyse the forecasting performance of business confidence survey indicators (*i.e.* Hansson *et al*., 2003, Lemmens *et al*., 2005; Abberger, 2006; Claveria *et al.*, 2007; Cesaroni *et al*., 2011; Frale and Monteforte, 2011; Girardi *et. al*., 2016). The ability of business survey indicators coming from the BoI survey on Inflation and Growth Expectations (SIGE) to forecast business cycle evolution and to lead turning points has been discussed in Cesaroni and Iezzi (2017), while the ability of Istat business surveys indicators to predict business cycle has also been analysed in Bruno and Lupi, (2004), Cesaroni (2011) and Bruno *et al*. (2019). In particular, Cesaroni and Iezzi (2017) use Harding and Pagan's (2002) methodology to detect the business cycle turning points of eight SIGE indicators[4]; they find that both the general economic situation and employment expectations indicators can lead their reference series (GDP and employment) turning points. Analogously, Bruno and Malgarini (2002) analyse Istat business confidence indicators turning points using the Bry-Boschan algorithm.

Business survey data were also used to build a measure of economic uncertainty by Bachmann *et al*. (2013), who introduced a measure of macroeconomic uncertainty given by the variance of a linear combination of positive and negative judgements reported from surveys. In their findings, positive innovations to sectoral uncertainty have prolonged negative implications for sectoral economic activity in the same way as adverse sectoral business confidence shocks.

The paper is structured as follows: Section 2 describes the time series considered in the analysis, as well as their use to track macroeconomic uncertainty. Paragraph 3 focusses on the possible sources of the differences between alternative business survey indicators. Paragraph 4 deals with the statistical characteristics of the Istat and SIGE surveys and provides a nowcasting exercise, while Paragraph 5 concludes this paper.

4 Namely, inflation expectations, expectations about firms own selling prices, employment expectations, three months investment conditions, three years expectations on investment conditions, expectations on the general economic situation, and probability of economy improvement in the next three months.

2. Data

The time series that we are going to analyse are indicators of the general economic situation (SITGEN) and firms' employment expectations (EMPL_EXP). These indicators were collected from both the BoI and Istat surveys. The BoI business survey indicators come from the Survey on Inflation and Growth Expectations (SIGE), which started in 1999, and the results are available on the Bank of Italy's website[5]. The Istat survey data come from the business survey on manufacturing firms, which started in the 1980s[6]; the time series are available on the Istat database website (I.Stat). The similarity of the indicators collected in these two Italian surveys calls for a comparison of their cyclical properties to detect the differences in their predictive ability.

Concerning the data frequency, while Istat collects these indicators on a monthly basis, the SIGE survey is conducted on a quarterly basis. In the Istat survey, firms are interviewed using mixed techniques, including Computer-Assisted Telephone Interviews (CATI) and responses via fax/mail. In the BoI survey, 5 per cent of firms were contacted by telephone and 95 per cent via the web.

Since Istat confidence indicators are available at a monthly frequency while BoI business confidence indicators are collected on a quarterly basis, to make a full and fair comparison in terms of predictive accuracy, we transformed the Istat indicators from monthly to quarterly frequency. In more detail, we constructed the quarterly time series of the Istat business survey indicators taking as representative of the quarter the value of the month in which the survey of the Bank of Italy indicator was collected.

In both business tendency surveys, respondents had three reply options for each question: up, same, down, or above normal, normal, and below normal. To convert the number of answers to each of the three reply options into percentages, the information was transformed into a balance given by the frequency of positive judgements (P) minus the frequency of negative judgements (N):

$$B = 100\ (P–N).$$

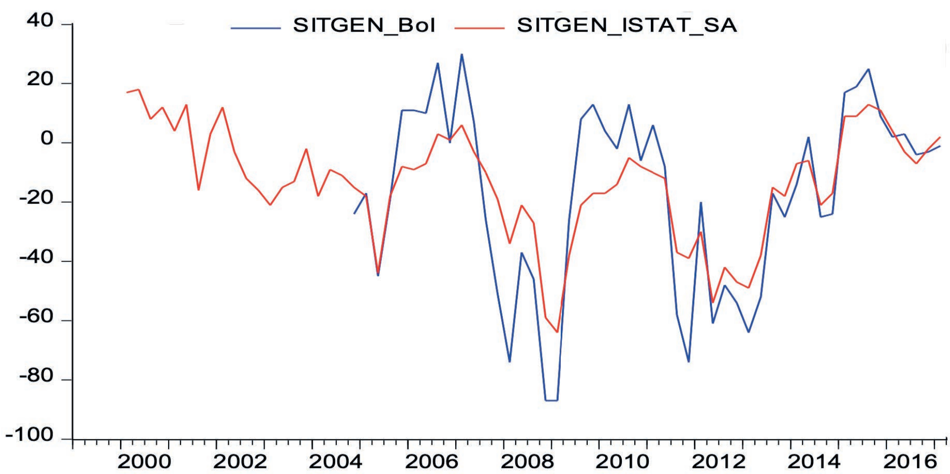
5 https://www.bancaditalia.it/pubblicazioni/indagine-inflazione/index.html?com.dotmarketing.htmlpage.language=1.

6 In Italy, manufacturing business survey data started to be collected by ISAE institutes during the 1980s within the joint harmonised programme of the European Commission. Currently, data are collected by Istat.

Balance is generally used as a proxy for business cycle evolution; if the number of positive judgements is greater than the number of negative judgements, we expect to approximate a positive business cycle phase.

We show, in Figure 2.1, the expectations on the general economic situation from BoI (SIT_GEN_BoI) and expectations on the general economic situation from Istat (SIT_GEN_ISTAT); Figure 2.2 also compares the employment expectations from BoI (EMP_EXP_BoI) and employment expectations from Istat seasonally adjusted (EMPL_EXP _ISTAT_sa). The sample considered spans from 2004Q4 (that is, the date on which SIT_GEN_BoI and EMP_EXP_BoI indicators started to be collected) to 2017Q1.

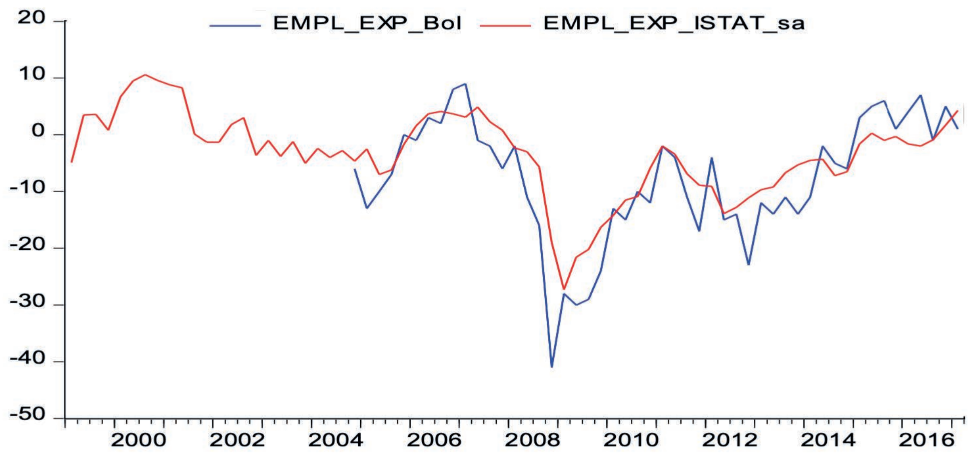
Figure 2.1 - The general economic situation expectations in the Istat and SIGE surveys. Years 2000-2016



Source: Authors' processing

As expected, the time series extracted by the two surveys are quite similar. However, there are some differences in the dynamics that can arise from various factors such as the sample, the exact timing in which the interview is conducted, the seasonal adjustment procedure and the questionnaire design.

Figure 2.2 - The employment expectations in the Istat and SIGE surveys. Years 2000-2016



Source: Authors' processing

2.1 Macroeconomic uncertainty and business confidence indicators

Often, information from business survey data is used to assess the uncertainty surrounding the economic environment. Indeed, their timeliness makes them suitable to evaluate changes in macroeconomic evolution and to track elements of uncertainty in the qualitative judgements directly coming from firms' management.

To compare and detect other possible differences in the uncertainty signal coming from BoI and Istat surveys, we also try to estimate a measure of the uncertainty on the macroeconomic environment, using the answers to the general economic situation question. More in detail, following Bachmann *et al.* (2013), we use an uncertainty measure based on the following formula:

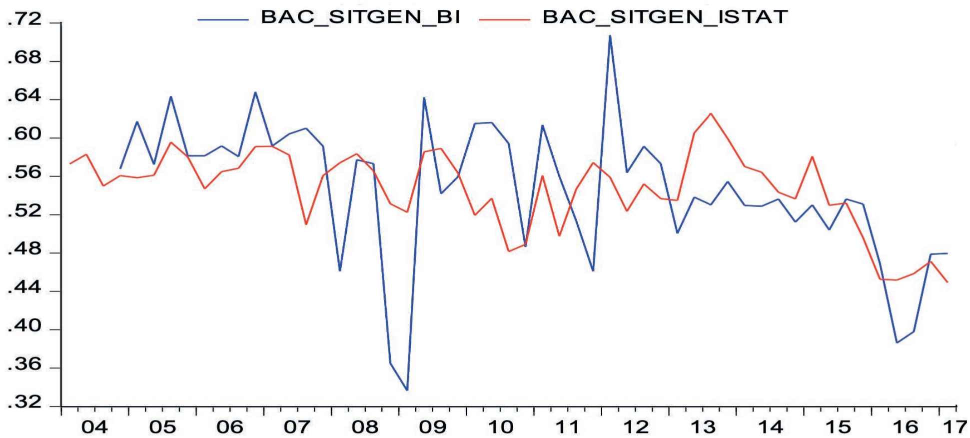
$$U_t = sqrt\{[frac_t(Increase) + frac_t(Decrease)] - [frac_t(Increase) - frac_t(Decrease)]\}^2$$

where U_t is the uncertainty given by the cross-sectional standard deviation of the survey responses. More in detail, $frac_t$ (*Increase*) represents the fraction of respondents that indicates an increase, quantified with +1, $frac_t$ (*Decrease*) represents the fraction of respondents that suggests a decrease of the survey indicators, and it is quantified with -1.

Figure 2.3 compares the Bachmann index calculated for the general economic situation question both in BoI and Istat surveys.

Looking at Figure 2.3, we see that the Bachmann (2013) index on macroeconomic uncertainty calculated using the SIGE data seems to display a higher variability concerning the corresponding Istat index.

Figure 2.3 - The Bachman index computed from BoI and Istat surveys. Sample 2004-2017



Source: Authors' processing

3. Possible sources of heterogeneity between the two surveys

In this paragraph, we focus on the sources of the differences between the indicators extracted from the two surveys, regarding the sample selection, the aggregation procedures of microdata, the seasonality treatment and the frequency.

3.1 The sample

The sample design and the structure of the sample selected can have a non-negligible influence on the results. There is heterogeneity in the population, concerning size or other characteristics; the use of stratification-based sampling methods is recommended. To give an idea of the differences in the sample structure of Istat and BoI business survey, Table 3.1 reports the stratification structure of the sample by economic sectors in both surveys. However, in this paper, we will only consider the survey conducted on the industrial sector, excluding construction firms, as this is the sector for which the two surveys (on the general economic situation and the employment expectations) are most consistent.

Table 3.1 Istat *vs*. Bank of Italy manufacturing firms' samples

	Istat sample	Bank of Italy sample
Industry excluding construction	4,100	410
Services	2,000	420
Construction	500	210

Source: Authors' processing

Looking at the Table, we can see that the Istat Survey on manufacturing firms is based on a sample of roughly 4,100 firms from the industrial sector interviewed each month. Furthermore, Istat conducts a survey on 2000 firms from the services sector and a survey on roughly 500 firms from the construction sector. The three surveys are undertaken separately, and there is no common stratification of firms among these three sectors[7]. Bank of Italy considers instead a smaller sample of roughly 1,000 firms interviewed on a quarterly basis, stratified on the basis of industry, services and construction

7 In the Istat manufacturing survey, firms are stratified according to value added and number of employees.

sectors. The main difference in the sample of the two surveys is that while Istat considers all firms with more than 5 employees, the firms considered in the Bank of Italy SIGE survey are larger, with more than 50 employees[8].

The differences in the two samples also depend on the importance of revenues of firms with more than 5 employees on the total revenues. To deeply inspect this aspect, using Istat survey microdata, it would be possible to construct a business confidence indicator taking only the larger firms, as the Bank of Italy survey does. In this way, the universe between the two indicators can be more comparable. However, since Istat microdata are not public, we leave this aspect for future research.

3.2 Seasonality treatment

Another difference between the two surveys analysed concerns the seasonal treatment of the data. The impact of seasonal adjustment in forecasting models has been widely studied. Proietti (2012) finds that the larger the seasonal components, the larger will be the estimation error for the seasonally adjusted series, which in turn will yield a less reliable cycle estimate. Fok *et al*. (2006) analyse the ability of TRAMO SEATS and CENSUS X-12 ARIMA algorithms to detect seasonality in the data and find that the methods seem to have a similar performance. More related to the empirical application of this paper, Mazzi and Savio (2005) investigated the possible existence of stochastic trends and seasonal unit roots in business tendency surveys. They concluded that the series were not affected by seasonality and thus they should not be "treated" to remove seasonality.

Istat considers a seasonally adjustment procedure for business survey data results based on a Tramo-Seats algorithm[9]. Quite the opposite, the Bank of Italy survey indicators coming from SIGE are not seasonally adjusted; in Section 4, using the Tramo-Seats algorithm, we also do not find evidence of seasonality in the time series of SIGE considered in this paper. Furthermore, the BoI survey considers raw data when computing statistics of SIGE to prevent cyclical frequency leakages due to the seasonal adjustment procedure.

8 The weighting procedure used to expand the survey results to the universe can also potentially have an impact on the aggregate results.

9 The standard seasonality procedures are included in the EC manual.

3.3 Question design

The questionnaire design, namely how a question is formulated, can also have an impact on the reliability of the answers and thus on the uncertainty surrounding the business survey indicators. As concerns the indicators here examined, there are some differences in the wording of the questions between the two business surveys. In SIT_GEN_BoI, the respondents are asked to assess Italy's general economic situation compared to three months ago, while in SIT_GEN_ISTAT, they are asked to evaluate the general tendency of the Italian economy, abstracting from sectoral and firms' developments in the next three months. For the question concerning employment EMPL_EXP, the wording of the questions is identical in the two surveys. In both of them, the respondents are asked to assess the number of employees in firms in the next three months. The exact formulation of the questions can be found in the Appendix.

Concerning the general economic situation indicator, while Istat asks the respondents to provide an expectation of the general economic situation in the next three months, abstracting from the situation of their own business, the BoI survey asks the respondents to judge the general economic situation in the current quarter, without specifying to disregard their business situation. This difference in the wording of the questions and of the reference period can induce a different information set on which firms formulate their expectations and provide a mismatch in the signal coming from the indicator across the two surveys.

However, the paper of Cesaroni and Iezzi (2017) on the ability of SIGE data to detect business cycle turning points and to forecast the business cycle showed that even if the general economic situation question is based on the current situation, the resulting indicator has leading properties on the business cycle. This may be because respondents are not able to correctly distinguish between the current and the future situation in the very short term. When looking at the present situation, firms may also consider elements concerning the very near future and vice versa, given the very short time horizon (3 months).

Quite the opposite, the formulation of the two questions about the employment expectation is virtually the same in both BoI and Istat surveys.

4. Empirical analyses

This section deals with the statistical properties of business survey time series considered in the forecasting application. Sub paragraph 4.1 explores the statistical properties of the two survey indicators considered (employment expectation, general economic situation), regarding the seasonality, the stochastic trends and the unconditional cross-correlations with their reference series, namely the industrial production, the industrial value added and the employment. Sub paragraph 4.2 reports the results of a forecast performance exercise, while sub paragraph 4.3 reports the predictive accuracy results.

4.1 Preliminary data analysis

The first characteristic of the survey indicators that we explore regards the eventual seasonality of SIGE indicators, given that they are reported in their raw form and are not treated. To detect possible seasonal patterns in these series, we report the results of the Tramo-Seats algorithm on SIT_GEN_BoI and EMP_EXP_BoI indicators.

Table 4.1 Estimated model by TRAMO-SEATS

	SIT_GEN_BoI	EMPL_EXP_BoI
SARIMA Model (P, D, Q) (p, d, q)	(1,0,0) (0,0,0)	(0,0,0) (0,0,0)

Source: Authors' processing

The first row of the table reports the SARIMA models (P, D, Q) (p, d, q) identified by TRAMO-SEATS, where (P, D, Q) are the order of AR component, of integration and Q and MA component of the non-seasonal model, while p, d, q are the corresponding parameters components of the seasonal component.

Looking at the results, we can notice that for the series SIT_GEN, there are no seasonal roots. Analogously, looking at the identified SARIMA model for employment expectations, we can see that there are no seasonally significant components identified by the algorithm as well.

Based on such empirical evidence, in what follows, we consider the not seasonally adjusted business survey series from the SIGE survey.

As preliminary data analysis, we also conducted a unit root test on both Istat and Bank of Italy indicators. Although the data analysed are expected to be stationary by construction, since they are built as a balance between positive and negative answers of respondents, in empirical samples, they might display a stochastic trend. This is because the upper and lower bounds for the values of those variables do not eliminate the possibility of local non-stationary data trends. The Augmented Dickey Fuller (ADF) tests for unit roots were implemented to identify possible non-stationary behaviours in the Business Survey variables.

Since the low power of the ADF test, a more powerful GLS test developed by Elliot, Rotemberg and Stock (1996) was also performed. The number of lags was chosen based on the Schwartz information criterion. The results are reported in Table 4.2.

Table 4.2 - Unit root tests (a)

	ADF	GLS
Sitgen_BI	-2.80(***)	-2.82(*)
Sitgen_SA_ISTAT	-2.49(***)	-2.27(*)
Sitgen_NSA_ISTAT	-2.55(***)	-2.41(**)
Empl_Exp_BI	-3.26(**)	-3.29(*)
Empl_exp_sa_ISTAT	-2.28(***)	-2.24
Empl_exp_nsa_ISTAT	-1.75(***)	-1.51(***)

Source: Authors' processing
(a) The sample goes from 2004 q1 to 2017q1; test with trend and intercept.
*: rejection of the unit root hypothesis at 1% level; **: rejection of the unit root hypothesis at 5% level; ***: rejection of the unit root hypothesis at 10% level.

We can reject the null hypothesis of a unit root; therefore, we treat the time series as stationary.

Finally, to analyse the link between survey indicators and the reference business cycle, the following Tables (4.3*a*, 4.3*b*, and 4.3*c*) report the cross-correlations with the reference business cycle indicators (industrial production, value-added and employment in the industrial sector) for the business survey data from BoI and Istat surveys. The value added is considered an alternative to the production index to track the business cycle of the industrial sector.

We did not consider the GDP as a reference indicator. However, it is usually the primary variable used to track the business cycle, since we focus on the survey of industrial firms.

As for survey indicators, we consider both the balance (difference between positive and negative answers provided by firms) and the components, namely the positive and the negative answers. The rationale for investigating the positive and negative answers in isolation is that the components of the balance may have a larger informative content than their difference, depending on the specific stage of the business cycle. For example, the negative (or positive) answer could have an ability to anticipate troughs (or peaks) or to display a different variability.

Finally, since BoI indicators are not seasonally adjusted, due to the absence of statistically relevant seasonal patterns, to have a fair comparison with Istat survey indicators, in the analysis, we consider both Istat seasonally and not seasonally adjusted data. The common sample considered is 2004q4 -2017q1.

Table 4.3*a* - Cross-correlations of survey indicators (balances) with their reference series

	t-4	t-3	t-2	t-1	t	t+1	t+2	t+3	t+4
					Industrial production yoy growth rate				
Sitgen_BI	0.46	0.58	0.70	0.65	0.46	0.22	-0.01	-0.18	-0.26
Sitgen_SA_ISTAT	0.18	0.40	0.61	0.69	0.62	0.46	0.25	0.06	-0.07
Sitgen_nsa_ISTAT	0.20	0.40	0.61	0.69	0.63	0.46	0.25	0.06	-0.08
					Industrial value added yoy growth rate				
Sitgen_BI	0.43	0.57	0.67	0.62	0.44	0.21	-0.02	-0.16	-0.25
Sitgen_SA_ISTAT	0.16	0.37	0.57	0.64	0.59	0.44	0.23	0.07	-0.07
Sitgen_nsa_ISTAT	0.16	0.37	0.57	0.64	0.59	0.44	0.23	0.06	-0.07
					Employment yoy growth rate				
Empl_exp_BI (a)	0.67	0.81	0.87	0.84	0.75	0.60	0.42	0.24	0.07
Empl_exp_sa_ISTAT	0.58	0.75	0.86	0.90	0.83	0.70	0.52	0.36	0.21
Empl_exp_nsa_ISTAT	0.56	0.72	0.82	0.83	0.77	0.64	0.50	0.36	0.24

Source: Authors' processing
(a) Sample from 2004q4.

Looking at the results reported in Table 4.1*a*, we can notice that, considering industrial production as the reference cycle, the general economic situation (SITGEN) indicators coming from Istat, both seasonally and not seasonally adjusted, show a higher contemporary correlation (0.62-0.63) than the BoI SITGEN indicator (0.46). Quite the opposite, the SIT_GEN_BoI indicator shows a higher correlation two quarters before showing a leading property.

Also considering the value added as reference business cycle, BoI balance (SITGEN_BoI) provides a higher correlation two quarters before therefore and in the sample analysed, seems to show a higher leading behaviour for the Istat indicator (SITGEN_I). Quite the opposite, the Istat indicator displays a higher contemporary correlation (0.59) than the BoI one (0.44).

An interesting result coming from this analysis concerns the ability of SITGEN_BoI to lead the business cycle, regardless of the wording of the question provided to firms, which does not explicitly state the expectation for the next three months, as in the analogous Istat question. One explanation for this finding can be related to the fact that, since in BoI the question on the general economic situation is not explicitly required to abstract from their business financial situation, this latter information is probably included in the information set that firms use to formulate their judgement on the General Economic Situation.

Looking at the cross-correlation between the employment cycle and the employment expectations (EMPL_EXP), we notice that the seasonally adjusted and not seasonally adjusted Istat indicators display a higher contemporary correlation (0.83 and 0.77, respectively) with respect to the corresponding BoI indicator (0.75). Both series are found to be leading the employment business cycle. However, while EMPL_EXP_BoI is leading two quarters ahead (0.87), EMPL_EXP_ISTAT is leading one quarter ahead.

Overall, we notice that seasonally adjustment of the Istat data does not seem to have a significant impact on the correlation properties with the reference cycle.

Tables 4.3*b*, 4.3*c* and 4.3*d* report the cross-correlations of positive, negative and stationarity judgements separately with real economic activity (industrial production and value added) and employment growth rate reference cycles for both SIGE and Istat surveys. For Istat, as for SIGE, we can only consider raw time series, as the Istat seasonal adjusted data are only available for the balances.

As expected, (Table 4.3b), the positive answers to the question on the general economic situation are procyclical and seem to have a leading power of two quarters concerning both the value added and the industrial production reference cycle. The correlation in t-2 is higher considering the BoI indicator. Quite the opposite, the Istat indicator seems to have a higher contemporary correlation with the BoI SITGEN indicator. Employment expectations judgments seem to have a leading power in t-2 that is similar in both survey indicators.

The negative judgements, both on the general economic situation and on the employment expectations (Table 4.3*c*), show a larger correlation with

Table 4.3*b* - Cross-correlations of survey indicators (positive judgments) with their reference cycle (a)

	t-4	t-3	t-2	t-1	t	t+1	t+2	t+3	t+4
					Industrial production yoy growth rate				
Sitgen_BI	0.43	0.49	0.56	0.50	0.35	0.17	-0.03	-0.17	-0.22
Sitgen_nsa_ISTAT	0.23	0.32	0.44	0.47	0.43	0.34	0.20	0.10	0.02
					Industrial value added yoy growth rate				
Sitgen_BI	0.38	0.47	0.54	0.48	0.34	0.16	-0.04	0.16	-0.21
Sitgen_nsa_ISTAT	0.17	0.28	0.40	0.43	0.40	0.32	0.19	0.11	0.04
					Employment yoy growth rate				
Occtot_BI	0.65	0.77	0.81	0.79	0.72	0.62	0.48	0.34	0.21
Empl_nsa_ISTAT	0.68	0.78	0.80	0.77	0.66	0.54	0.42	0.29	0.19

Source: Authors' processing

(a) Sample from 2004 q4 to 2017q1; not seasonally adjusted data.

Table 4.3*c* - Cross-correlations of survey indicators (negative judgments) with industrial production growth (yearly) (a)

	t-4	t-3	t-2	t-1	t	t+1	t+2	t+3	t+4
					Industrial production yoy growth rate				
Sitgen_BI	-0.45	-0.58	-0.71	-0.67	-0.48	-0.23	0.01	0.17	0.26
Sitgen_ISTAT	-0.17	-0.40	-0.63	-0.72	-0.66	-0.48	-0.25	-0.04	0.11
					Industrial value added yoy growth rate				
Sitgen_BI	-0.42	-0.57	-0.68	-0.63	-0.45	-0.22	0.01	0.15	0.25
Sitgen_ISTAT	-0.14	-0.38	-0.60	-0.68	-0.62	-0.45	-0.23	-0.05	0.11
					Employment yoy growth rate				
Empl_exp_BI	-0.50	-0.62	-0.68	-0.65	-0.56	-0.42	-0.25	-0.09	0.08
Empl_exp_nsa_ISTAT	-0.42	-0.61	-0.73	-0.77	-0.74	-0.63	-0.50	-0.37	-0.24

Source: Authors' processing

(a) Sample from 2004 q4 to 2017q1; not seasonally adjusted data.

the reference cycle concerning the positive judgments and, in some cases, also for the balances. SITGEN_BoI is confirmed, especially leading in two quarters for both industrial production (-0.71) and value added (-0.68) reference cycles. SITGEN_ISTAT seems instead to be leading one quarter ahead concerning both industrial production (-0.72) and value added (-0.68).

Looking at the correlation with the employment business cycle, EMPL_EXP from Istat displays the highest correlation with the employment growth rate in t-1 (-0.77). Quite the opposite, EMPL_EXP from SIGE displays the highest correlation in t-2 (-0.68).

To assess the possible impact of the time aggregation method on the cyclical properties of the data, the following table reports a sensitivity analysis of cross-correlations concerning time aggregation from monthly to quarterly frequency. Indeed, as explained before, the original frequency of Istat survey data is monthly while that of BoI indicators is quarterly; therefore, for the Istat quarterly data, we consider both the value of the month in which the BoI survey is carried out (*Sitgen_nsa_ISTAT and Empl_exp_nsa_ISTAT*) and the quarterly average of three months (*Sitgen_nsa_ISTAT_mean, Empl_exp_nsa_ ISTAT_mean*).

Table 4.4 - Cross-correlations of survey indicators with the reference variable (yearly). Sensitivity analysis concerning the time aggregation method (a)

	t-4	t-3	t-2	t-1	t	t+1	t+2	t+3	t+4
					Industrial production yoy growth rate				
Sitgen_BI	0.46	0.58	0.70	0.65	0.46	0.22	-0.01	-0.18	-0.26
Sitgen_nsa_ISTAT_mean	0.13	0.33	0.53	0.67	0.71	0.61	0.40	0.19	-0.03
Sitgen_nsa_ISTAT	0.20	0.40	0.61	0.69	0.63	0.46	0.25	0.06	-0.08
					Employment yoy growth rate				
Empl_exp_BI*	0.67	0.81	0.87	0.84	0.75	0.60	0.42	0.24	0.07
Empl_exp_nsa_ISTAT_mean	0.50	0.69	0.82	0.87	0.84	0.72	0.58	0.43	0.28
Empl_exp_nsa_ISTAT	0.56	0.72	0.82	0.83	0.77	0.64	0.50	0.36	0.24

Source: Authors' processing

(a) Sample from 2004 q4 to 2017q1; not seasonally adjusted data.

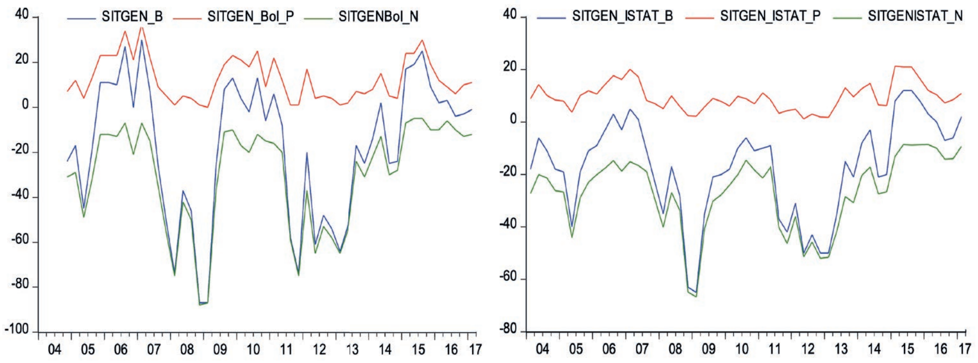
Looking at the results reported in the table, we can notice that the Istat indicator built considering the last month of the quarter (*Sitgen_nsa_ISTAT*), which is entirely consistent in timing for the alternative indicator in the SIGE survey, has a larger leading correlation concerning the indicator computed as average of the three months (variables with the suffix_mean).

Given the focus on nowcasting and forecasting of the empirical application, we choose to focus on the best indicator (taking the data of the last month of the quarter) for the Isat survey, also to have a fair comparison in terms of the informative set with SIGE[10].

To further inspect the linkages between balance, positive and negative answers dynamics, figures 4.1 and 4.2 report a comparison of positive, negative answers and the balance of the general situation and employment expectations of BoI and Istat surveys.

10 Since the BoI indicators consider only the xt information set, using a time aggregation based on xt, xt-1 and xt-2 for Istat, could penalise BoI in the statistical comparison.

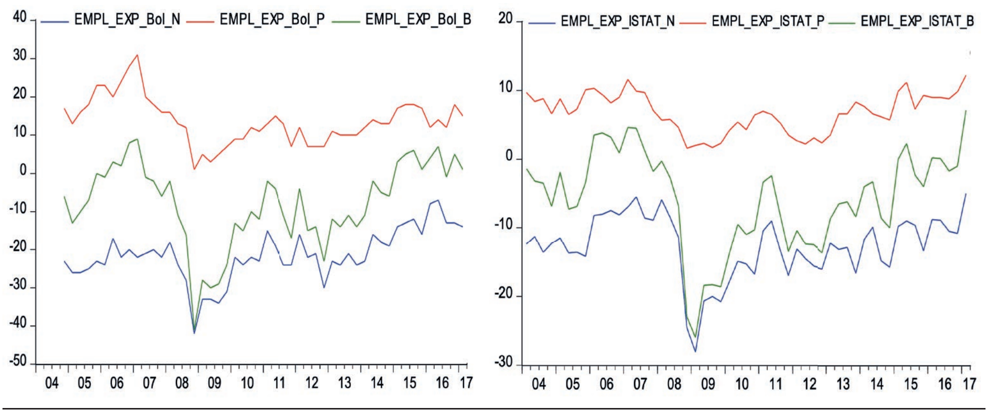
Figure 4.1 - General economic situation: positive, negative responses and the balance in the BoI and the Istat surveys. Sample 2004-2017



Source: Authors' processing

Looking at the dynamics of SITGEN negative judgements, we can see that the dynamics of the overall index are mainly explained by the negative opinion component, for both the general economic situation and for employment. More specifically, during the 2009 recession, they were more negative among BoI respondents for the Istat survey; this is also true when looking at negative judgements concerning the 2012 recession. Looking at the dynamics of SITGEN positive judgements, we can also notice that between 2005 and 2007, the BoI percentage of positive answers was higher than that provided by the Istat sample firms.

Figure 4.2 - Figure 4.2 Employment expectations: positive, negative responses and the balance in the BoI and the Istat surveys. Sample 2004-2017



Source: Authors' processing

Looking at the dynamics of the employment expectations, we see (Figure 4.2) that the negative judgements are usually lower in SIGE compared to the Istat survey. Concerning the dynamics of positive judgements, we can notice that for the general economic situation question, a higher percentage of positive judgements occurred during the biennium 2005-2007. Concerning the BoI SITGEN indicator, we can see that during the years 2007-2009, the dynamics of the balances were entirely driven by the negative answers provided by firms, while the positive judgements were roughly near zero. Again, the negative judgements are structurally further down in the SIGE survey.

Overall, the results show that the cyclical dynamics of the balances are mainly driven by the negative judgments, for both the indicators considered; the negative judgments in SIGE tend to be lower than in the Istat survey. Another information coming from qualitative indicators concerns stability judgements provided by firms in business cycle analysis. However, since this variable is more related to the trend evolution of the economy and more volatile during recession episodes and since we concentrate on short-term analysis concerning business cycle evolution, in what follows, we only analyse the predictive content of the balances, the positive and the negative judgements.

4.2 The forecast performance exercise

In what follows, we use an unrestricted bivariate VAR model to compare the forecasting power of BoI and Istat business survey indicators, which takes the form:

$$\begin{bmatrix} \boldsymbol{x_t} \\ \boldsymbol{y_t} \end{bmatrix} = \boldsymbol{\alpha_0} + \boldsymbol{\alpha_1} \begin{bmatrix} \boldsymbol{x_{t-1}} \\ \boldsymbol{y_{t-1}} \end{bmatrix} + \boldsymbol{\alpha_2} \begin{bmatrix} \boldsymbol{x_{t-2}} \\ \boldsymbol{y_{t-2}} \end{bmatrix} + \ldots\ldots + \boldsymbol{\alpha_n} \begin{bmatrix} \boldsymbol{x_{t-n}} \\ \boldsymbol{y_{t-n}} \end{bmatrix}$$

where $\mathbf{x}_t$ is the macroeconomic target (*i.e.* industrial production or employment) and $\mathbf{y}_t$ is its business survey indicator (namely BoI or Istat indicators). Hence, we consider two bivariate unrestricted VAR models for each survey. More in detail, we compare and analyse the out-of-sample forecast performance of SITGEN and EMPL_EXP indicators concerning their corresponding aggregate series, namely industrial production (and value added) and employment, namely the number of persons employed in the manufacturing sector. To achieve this end, we use recursive estimates.

Since the dependent variables display a stochastic trend and survey indicators are stationary, we forecast variables in year-on-year differences to remove the unit root in the time series. Given that the data are quarterly, we use the fourth differences (in logarithms). According to this notation in the model, $xt=\Delta^4 logXt=(1-L^4)logXt$, where L is the lag operator and 4 is the order of differencing.

The following Tables report the results of a forecast exercise conducted considering an estimation sample from 2004 q4 to 2012 q4 and a forecasting window equal to 16 quarters from 2013 q1 to 2016 q4. In the exercise, we compute the RMSFE in a dynamic *ex ante* forecast setting, considering recursive estimation methods.

An automatic parsimonious lag pre-selection based on a general-to-specific algorithm is used to specify the functional form of the models (see Hendry and Krolzig, 2001). More in detail, starting from an initial 6 lag model, the algorithm selects reduced form models based on lag statistical significance.

The recursive window for the out-of-sample exercise spans from 19 quarters for 1 1-step ahead forecast to 14 quarters for 4-step ahead forecasts. The lags of explanatory variables reported in the following tables are statistically significant.

Table 4.4 reports the results of 3 vector autoregressive models to forecast industrial production. Model 1 uses SIT_GEN_BoI. Models 2 and 3 use SITGEN_ISTAT both seasonally and not seasonally adjusted.

Table 4.4 - Forecasting models for the industrial production using SITGEN (balances) RMSFE for dynamic forecasts (a)

	Model 1 SITGEN_BI	Model 2 SITGEN_ISTAT_SA	Model 3 SITGEN_ISTAT_NSA
Intercept	0.012(***)	0.021	0.022
Δ^4IPI t-1	0.68(***)	0.825(***)	0.83 (***)
Δ^4IPI t-3		-0.37(**)	-0.38 (***)
SIT_GEN t-1	0.001(***)	0.0015(**)	0.0014 (**)
SIT_GEN t-4		0.0001	0.0002
Normality test (residuals)	4.81 (0.09)	4.95 (0.08)	4.48 (0.11)
Heteroscedasticity test	4.06 (0.01)	1.41 (0.25)	1.00 (0.46)
AR test (residuals)	3.03 (0.05)	0.31 (0.82)	0.28 (0.84)
1-step dynamic forecast	0.019505	0.022626	0.020825
2-step dynamic forecast	0.016088	0.018894	0.021251
3-step dynamic forecast	0.019670	0.021406	0.019025
4-step dynamic forecast	0.026446	0.019970	0.022635

Source: Authors' processing

(a) The dependent variable is the year-on-year growth rate of industrial production (IPI). Estimation sample: 2005 Q1-2012 Q4. Forecasting sample: 2013 Q1-2016Q4.
*: significant at 10 per cent; **: significant at 5 per cent; ***: significant at 1 per cent; p-values in parenthesis.

The models of Table 4.4, estimated using the indicators on the general economic situation show that in terms of RMSFE model 1, based on SIGE, seems to perform better than the alternatives for 1 and 2 steps ahead; model 3 and model 2, both based on the Istat survey, are respectively better for 3 and 4 steps ahead. The diagnostics indicate no heteroscedasticity and normality for models 2 and 3. The AR test on the autocorrelation of the residuals, set on lag length from 1 to 4, shows that model 1 has no autocorrelation, models 2 and 3 have autocorrelation. The selected model 1 displays a degree of autocorrelation and heteroscedasticity in the residuals, while for the normality test, the significance is on borderline thresholds. However, the model has been selected by the general to specific algorithm, and the residuals' behaviour does not invalidate its forecast performance accuracy, but potentially only the coefficients' interpretation, which is not the focus of the paper.

We also need to inspect if the tail indicators on the general economic situation (*i.e.* positive and negative judgements) considered individually have a forecasting content. Table 4.5 shows the models to forecast industrial production using SITGEN_BoI positive judgements, and Table 4.6 focusses on the negative judgements.

Table 4.5 - Forecasting models for the industrial production using SITGEN (positive) RMSFE for dynamic forecasts

	Model 1 SITGEN_BI_P	Model 2 SITGEN_ISTAT_NSA_P
Intercept	-0.025**	-0.0165**
Δ^4IPI t-1	0.9***	1.283***
Δ^4 IPI t-2		-0.61***
Δ^4 IPI t-3	-0.31***	
SIT_GEN t-1	0.0014***	0.0018**
Normality test (residuals)	1.74 (0.42)	0.17 (0.92)
Heteroschedasticity test	0.84 (0.55)	0.62 (0.71)
AR test (residuals)	1.1 (0.37)	2.85 (0.06)
1-step dynamic forecast	0.024836	0.023498
2-steps dynamic forecast	0.023724	0.022595
3-steps dynamic forecast	0.024985	0.022485
4-steps dynamic forecast	0.027356	0.022920

Source: Authors' processing
(a) The dependent variable is the year-on-year growth rate of industrial production. Estimation sample: 2005 Q2-2012. Forecasting sample: 2013 Q1-2016 Q4.
*: significant at 10 per cent; **: significant at 5 per cent; ***: significant at 1 per cent; p-values in parenthesis.

Table 4.6 - Forecasting models for the industrial production using SITGEN (negative) RMSFE for dynamic forecasts

	Model 1 SITGEN_BI_N	Model 2 SITGEN_ISTAT_NSA_N
Intercept	0.016**	0.051
Δ^4IPI t-1	0.87***	0.786***
Δ^4 IPI t-3	-0.29***	-0.336**
SIT_GEN t-1	-0.00065**	-0.00212**
Normality test (residuals)	5.80 (0.05)	5.44 (0.06)
Heteroschedasticity test	0.99 (0.45)	0.31 (0.92)
AR test (residuals)	1.24 (0.32)	0.38 (0.76)
1-step dynamic forecast	0.025746	0.025898
2-steps dynamic forecast	0.023825	0.023562
3-steps dynamic forecast	0.025775	0.025945
4-steps dynamic forecast	0.025945	0.024879

Source: Authors' processing
(a) The dependent variable is the year-on-year growth rate of industrial production. Estimation sample: 2005 Q2-2012 Q4. Forecasting sample: 2013 Q1-2016 Q4.
*: significant at 10 per cent; **: significant at 5 per cent; ***: significant at 1 per cent; p-values in parenthesis.

Looking at the results of Table 4.5 using positive judgements on the general economic situation, we can notice that model 2 from Istat data performs slightly better than model 1 using SIGE in terms of RMSFE. Results using the negative judgements, reported in Table 4.6, show instead that model 1 in SIGE perform similarly to model 2, with marginal differences at each time horizon. All the models pass the residuals diagnostics.

In order to analyse the possible sensitivity with respect to the reference business cycle (namely industrial production *versus* value added), Table 4.7 reports the results of vector autoregressive models to forecast the manufacturing value added (instead of industrial production) by means of the general economic situation. Model 1 uses the balance of the general economic situation from BoI survey (SIT_GEN_BoI). Models 2 and 3 use SITGEN_ISTAT both seasonally and not seasonally adjusted.

Table 4.7 - Forecasting models for the industrial value added using SITGEN (balances). RMSFE for dynamic forecasts (a)

	Model 1 SITGEN_BoI	Model 2 SITGEN_ISTAT_SA	Model 3 SITGEN_ISTAT_NSA
Intercept	0.0083**	0.0111	0.0109
Δ^4VA t-1	1.26***	1.292***	1.30***
Δ^4VA t-2	-0.59***	-0.63***	
Δ^4VA t-3			-0.63***
SITGEN t-1	0.00033***	0.00041	0.00036
Normality test (residuals)	2.31 (0.31)	1.86 (0.39)	2.36 (0.31)
Heteroschedasticity test	1.31 (0.29)	0.97 (0.46)	1.04 (0.42)
AR test (residuals)	0.70 (0.57)	0.31(0.81)	0.38 (0.77)
1-step dynamic forecast	0.01778	0.015306	0.014296
2-steps dynamic forecast	0.017566	0.018794	0.019941
3-steps dynamic forecast	0.017673	0.015837	0.014954
4-steps dynamic forecast	0.020056	0.018597	0.01771

Source: Authors' processing
(a) The dependent variable is the year-on-year growth rate of the value added in the industrial sector (VA). Estimation sample: 2005 Q2-2012 Q4 Forecasting sample: 2013 Q1-2016 Q4.
*: significant at 10 per cent; **: significant at 5 per cent; ***: significant at 1 per cent; p-values in parenthesis.

Results of forecasting models reported in Table 4.7 show that while models 2 and 3 from Istat perform better 1, 3 and 4 steps ahead, model 1 from SIGE performs better in terms of RMSFE for two quarters ahead. All the models pass the usual residuals diagnostics.

Tables 4.8 and 4.9 report the results of the VAR models to forecast manufacturing value added using, respectively, the positive and the negative judgements concerning the general economic situation.

Table 4.8 - Forecasting models for the industrial value added using SITGEN (positive). RMSFE for dynamic forecasts (a)

	Model 1 SITGEN_BI_P	Model 2 SITGEN_ISTAT_NSA_P
Intercept	-0.0094**	-0.0099
Δ^4VA t-1	1.28***	1.33***
Δ^4 VA t-2	-0.60***	-0.65***
SITGEN t-1	0.54**	0.0012
Normality test (residuals)	1.61 (0.44)	2.63 (0.26)
Heteroscedasticity test	0.91 (0.50)	0.96 (0.47)
AR test (residuals)	1.18 (0.34)	0.69 (0.57)
1-step dynamic forecast	0.013847	0.014296
2-step dynamic forecast	0.018921	0.019941
3-step dynamic forecast	0.016863	0.014954
4-step dynamic forecast	0.02225	0.017710

Source: Authors' processing
(a) The dependent variable is the year-on-year growth rate of the value added in the industrial sector (VA). Estimation sample: 2005 Q2-2012 Q4. Forecasting sample: 2013 Q1-2016 Q4.
*: significant at 10 per cent; **: significant at 5 per cent; ***: significant at 1 per cent; p-values in parenthesis.

Table 4.9 - Forecasting models for the industrial value added using SITGEN (negative). RMSFE for dynamic forecasts (a)

	Model 1 SITGEN_BoI_N	Model 2 SITGEN_ISTAT_NSA_N
Intercept	0.024	0.0016
Δ^4 VA t-1	0.83***	1.36***
Δ^4 VA t-2		-0.66***
Δ^4 VA t-3	-0.32***	
SITGEN t-2	-0.00074**	
Normality test (residuals)	4.43 (0.11)	1.48 (0.48)
Heteroscedasticity test	0.61 (0.71)	1.19 (0.34)
AR test (residuals)	0.17 (0.91)	0.014 (0.99)
1-step dynamic forecast	0.016995	0.013669
2-step dynamic forecast	0.015524	0.023404
3-step dynamic forecast	0.017301	0.019034
4-step dynamic forecast	0.022018	0.019933

Source: Authors' processing

(a) The dependent variable is the year-on-year growth rate of the value added in the industrial sector (VA). Estimation sample is: 2005 Q2-2012 Q4. Forecasting sample: 2013 Q1-2016 Q4.

*: significant at 10 per cent; **: significant at 5 per cent; ***: significant at 1 per cent; p-values in parenthesis

Looking at the results of Table 4.8 using positive judgements, we can notice that the BoI model performs better in the short term (one and two steps ahead), while model 2 from Istat performs better three and four steps ahead. Quite the opposite, the results of Table 4.9 considering negative judgements on the general economic situation show that the Istat model performs better than the BoI model at 1 and 4 steps ahead. The models pass the residuals diagnostics.

Overall, the results concerning the predictive ability of the general economic situation seem to show that there is no clear pattern. Depending on the time horizon or the reference variable, one indicator may dominate over the others, but still, the differences in the RMSFE are always minor.

We now focus on the predictions for the labour market. Table 4.10 reports the results of 3 vector autoregressive models to forecast employment; Model 1 uses EMPL_EXP_BoI, while models 2 and 3 use EMPL_EXP_ISTAT, respectively, seasonally and not seasonally adjusted.

Table 4.10 - Forecasting models for employment using employment expectations (balances). RMSFE for dynamic forecasts (a)

	Model 1 EMPL_EXP_BI	Model 2 EMPL_EXP_ISTAT_SA	Model 3 EMPL_EXP_ISTAT_NSA
Intercept	-0.0033 **	-0.0028 **	-0.0027
Δ^4 empllog t-1	1.044 ***	1.052 ***	1.053 ***
EMPL_EXP t-1	-0.00016	-0.00017 **	-0.00014
Normality test (residuals)	1.42 (0.49)	0.74 (0.69)	0.96 (0.61)
Heteroscedasticity test	0.22 (0.92)	0.29 (0.88)	0.38 (0.81)
AR test (residuals)	1.37 (0.27)	1.86 (0.16)	1.82 (0.17)
1-step dynamic forecast	0.029146	0.032682	0.055628
2-step dynamic forecast	0.04388	0.040978	0.068691
3-step dynamic forecast	0.02764	0.042765	0.072859
4-step dynamic forecast	0.01536	0.034263	0.020924

Source: Authors' processing
(a) The dependent variable is the year-on-year growth rate of employment. Estimation sample: 2005 Q2-2012 Q4. Forecasting sample: 2013 Q1-2016 Q4. p-values in parentheses
*: significant at 10 per cent; **: significant at 5 per cent; ***: significant at 1 per cent; p-values in parenthesis.

Looking at the RMSFE results of Table 4.10 at various steps, one can notice that the EMPL_EXP_BoI indicator performs better at 1, 3 and 4 steps ahead, while the RMSFE of the model with the series EMPL_EXP_ISTAT_SA is lower than the alternatives 2 steps ahead. The residual test shows normality, no heteroscedasticity and no autocorrelation.

The results of the models using the positive and negative judgements on employment expectations are respectively shown in Tables 4.11 and 4.12.

Table 4.11 - Forecasting models for employment using employment expectations (positive). RMSFE for dynamic forecasts (a)

	Model 1 EMPL_EXP_BI_P	Model 2 EMPL_EXP_IS_NSA_P
Intercept	0.00825	0.0014
Δ^4 empllog t-1	1.11***	1.46***
Δ^4 empllog t-2		-0.49**
EMPL_EXP t-1	-0.00043**	-0.000
Normality test (residuals)	1.61 (0.44)	3.01(0.23)
Heteroscedasticity test	0.42 (0.82)	0.36 (0.92)
AR test (residuals)	1.31 (0.29)	0.61 (0.61)
1-step dynamic forecast	0.037624	0.087372
2-step dynamic forecast	0.03109	0.1277
3-step dynamic forecast	0.01124	0.22736
4-step dynamic forecast	0.01989	0.022375

Source: Authors' processing
(a) The dependent variable is the year-on-year growth rate of employment. Estimation sample: 2005 Q2-2012 Q4. Forecasting sample: 2013 Q1-2016 Q4. p-values in parentheses.
*: significant at 10 per cent; **: significant at 5 per cent; ***: significant at 1 per cent; p-values in parenthesis.

Table 4.12 - Forecasting models for employment using employment expectations (negative). RMSFE for dynamic forecasts (a)

	Model 1 EMPL_EXP_BoI_N	Model 2 EMPL_EXP_ISTAT_NSA_N
Intercept	-0.01189**	-0.0047
Δ^4 empllog t-1	0.9758*	1.0579***
EMPL_EXP t-1	0.00042***	0.00045**
Normality test (residuals)	3.04 (0.21)	2.75 (0.25)
Heteroscedasticity test	2.27 (0.09)	2.75 (0.98)
AR test (residuals)	0.36 (0.78)	0.11(0.17)
1-step dynamic forecast	0.021144	0.053681
2-step dynamic forecast	0.052690	0.062785
3-step dynamic forecast	0.036270	0.074971
4-step dynamic forecast	0.014980	0.014551

Source: Authors' processing
(a) The dependent variable is the year-on-year growth rate of employment. p-values in parentheses. Estimation sample: 2005 Q2-2012 Q4. Forecasting sample: 2013 Q1-2016 Q4.
*: significant at 10 per cent; **: significant at 5 per cent; ***: significant at 1 per cent; p-values in parenthesis.

Looking at the RMSFE reported in Table 4.11, we can see that model 1 using BoI employment expectations positive judgements (EMPL_EXP_BI_P) performs better at all time horizons concerning the Istat indicator. The models pass the usual residuals diagnostics. Similarly, looking at the results in terms of RMSFE reported in Table 4.12 shows that model 1 using BoI negative judgements on the general economic situation (SITGEN) performs better than model 2 using Istat data at all time horizons. Model 2 passes all the residual diagnostics at a 1% significance level, while Model 1 passes the normality test and AR test at a 1% level and the heteroscedasticity test at a 10% level.

Overall, looking at the results of the forecast models for industrial production, value added and employment, we can notice that Bank of Italy SIGE survey indicators, although based on a small sample of firms, seem to display a forecast performance that is broadly similar to the predictions of the industrial output. At the same time, they tend to be better for SIGE regarding employment. The latter result might seem at odds with the fact that the Istat survey is based on a larger sample; one explanation for this result can be linked to the inclusion in the Istat sample of the small firms, which could bring some noise in the answers to the questionnaire. Larger firms probably formulate more reliable expectations and judgements, putting more resources into developing their plans concerning small firms; moreover, the person answering the survey questionnaire in the small enterprise may have less macroeconomic expertise than the person responding in the large enterprise. This interpretation concerning a limited ability of small firms to track and forecast business cycle evolution is coherent with Van Nieuwerburgh and Veldkamp's findings (2006). Another result concerns the fact that the seasonal adjustment does not seem to be so relevant, given that often the forecast accuracy of the models based on the raw Istat indicator is better than that of the models using the Istat seasonal adjusted data.

4.3 Comparing predictive accuracy

To assess if the difference in the forecasting performance among models is statistically significant, Diebold Mariano (1995) test results are also reported. The test is based on the null hypothesis that the forecast performance between two models is equal, against the alternative that it is statistically different. The test statistics is:

$$T = \frac{\bar{d}_t}{\left(cov\left(d_{t,}d_{t-1,}\right)/T\right)^{1/2}}$$

where $\bar{d}_t = \left(e^{BoI}_{t+h/t} - e^{ISTAT}_{t+h/t}\right)$ indicates the difference between the forecasting prediction errors obtained using the BoI and Istat forecasting models and $\bar{d}_t = \frac{1}{t_0}\sum_{t_0}^{T} d_t$.

Table 4.13 reports the p-values of the Diebold Mariano test for 1, 2, 3, and 4 forecast horizons for 2 bivariate VAR models used to forecast industrial production and employment.

Table 4.13 – Results of the Diebold Mariano test (p-values)

Test of equal accuracy of the Istat forecast wrt the BoI forecast
H0: Forecast accuracy is equal

	Industrial production		Employment	
	SITGEN_BOI SITGEN_ISTAT_NSA	SITGEN_BOI SITGEN_ISTAT_SA	EMPL_EXP_BoI EMPL_EXP_ISTAT_NSA	EMPL_EXP_BoI EMPL_EXP_ISTAT_SA
1-step forecast	0.03	0.87	0.00	0.02
2-step forecast	0.87	0.38	0.00	0.15
3-step forecast	0.13	0.08	0.00	0.37
4-step forecast	0.25	0.28	0.09	0.01

Source: Bank of Italy, Survey on Inflation and Growth Expectations (SIGE), and Istat, Manufacturing business survey

The results of the Diebold Mariano test indicate that the differences in the MSFE of the VAR models based on SIGE and Istat business surveys for industrial production are, in general, not statistically significant; this is consistent with the fact that the differences between the RMSFE are minor, as noted above. In the case of the employment prediction only, there is a dominance of the SIGE indicator over the raw indicator of Istat, as already mentioned, looking at the size of the prediction errors.

Based on the previous results, it seems that both surveys, SIGE and Istat, have similar informative content, in nowcasting and short-term forecasting. It is therefore worth investigating if one model may encompass an alternative one, in the sense that its forecast captures completely the information of the alternative.

In what follows, to test the relevance of business survey indicators information, the results of an encompassing test are also shown. The idea of the encompassing test relates to the notion that one model not only fully explains what another model can explain but also provides additional information. With competing forecasts, the condition that *f1,t+1/t* encompasses *f2,t+1/t* can be stated as:

$$E\left[L\left(f/_{1,t+1/}t, y_{t+1}\right)\right] \leq min_{g(.)} E\left[L\left(f/_{1,t+1/}t, f/_{2,t+1/}t\right), y_{t+1}\right]$$

If this equation holds, then the first forecast is sufficient in the sense that there is no information in the second forecast that is useful once we have access to the first forecast[11].

Table 4.14 reports the forecast encompassing tests based on two statistics, “test a” and “test b”. In “test a”, the null hypothesis is that the prediction based on the SIGE survey encompasses the one based on the Istat survey. In “test b”, the null hypothesis is that the prediction based on the Istat survey encompasses the one based on the SIGE survey. More in detail, the encompassing test is based on Chong and Hendry (1986), considering only forecast information and abstracting from the data-generating process and the model structure.

11 Note that the forecast encompassing depends on a standard quadratic loss function. However, as shown in Harvey et. al. (1998) the forecast evaluation results are not significatively sensitive to different loss functions.

Table 4.14 – Results of the forecast encompassing tests (p-values) (a)

		Industrial production		Employment	
		SITGEN_BOI SITGEN_ISTAT_NSA	SITGEN_BOI SITGEN_ISTAT_SA	EMPL_EXP_BoI EMPL_EXP_ISTAT_NSA	EMPL_EXP_BoI EMPL_EXP_ISTAT_SA
h=1	Test a	0.00	0.00	0.00	0.00
	Test b	0.00	0.00	0.00	0.00
h=2	Test a	0.00	0.00	0.00	0.00
	Test b	0.00	0.01	0.00	0.00
h=3	Test a	0.00	0.00	0.34	0.00
	Test b	0.00	0.00	0.00	0.00
h=4	Test a	0.04	0.06	0.02	0.40
	Test b	0.01	0.04	0.00	0.00

Source: Authors' processing

(a) Test a: the null hypothesis is that the prediction based on the SIGE survey encompasses the one based on the Istat survey; test F.
Test b: the null hypothesis is that the prediction based on the Istat survey encompasses the one based on the SIGE survey; test F.

The forecast encompassing test shows that, except for a few exceptions, the information contained in the SIGE survey may not encompass that of the Istat; therefore, both surveys are beneficial in nowcasting and short-term forecasting.

5. Conclusions and further research

In this paper, we analyse the similarities and differences between two business confidence indicators for Italy, on the expectations of the general economic situation and employment, collected by the Bank of Italy (Survey SIGE) and by Istat. To do this, we analyse the statistical properties of such data (*i.e.* seasonal components, unit roots and the cross-correlations concerning their reference cycles) and perform a forecast exercise using VAR models.

We find that the SITGEN indicator from the SIGE survey concerns its ability to lead the business cycle, similarly to the corresponding series in the Istat survey; this is not a foregone result, as the SIGE question to firms is not explicitly formulated as forward-looking, and the SIGE sample has a smaller number of firms than the Istat sample.

Regarding the predictions on employment, the SIGE indicator seems to overperform the Istat survey; the better forecast accuracy is also statistically significant for some horizons, but the predictions of the model based on SIGE do not encompass the alternative.

The forecasting exercise is mainly based on “balance”, which is given by the difference between positive and negative judgments provided by firms. We also explore the predictive power of its components, namely the positive and negative judgments considered separately. We find the predictive power of the positive and negative judgments is comparable to that of balance; therefore, they can be used singularly in nowcasting and short-term forecasting. This result holds for both the BoI and the Istat indicators.

Our results provide insights for further research. To better understand the extent to which the differences in the sample structure of the SIGE and manufacturing Istat surveys affect their leading properties, an in-depth study of the relevance of small firms in business surveys can be performed. Future studies could also explore forecasting combination techniques to take full advantage of the informative content of both surveys.

Appendix

Table A1 - Istat and Bank of Italy business survey questions

Bank of Italy	SITGEN_BoI	Compared with 3 months ago, do you consider Italy's general economic situation to be: 1 Better, 2 the same, 3 Worse?
Istat	SITGEN_Istat	In the next 3 months, the general tendency of the Italian economy, abstracting from sectoral and firms' developments, will be: 1 Favourable, 2 Stationary, 3 Unfavourable.
Bank of Italy	EMPL_EXP_BoI	In the next 3 months, overall employment of your firm will be: 1 Higher, 2 Unchanged, 3 Lower.
Istat	EMPL_EXP_Istat	Your firm's total number of employees in the next 3 months will be: 1 Lower, 2 Unchanged, 3 Higher.

Source: Authors' processing

References

Abberger, K. 2006. "Qualitative Business Surveys in Manufacturing and Industrial Production - What can be Learned from Industry Branch Results?". *Ifo Working Paper*, N. 31. München, Germany: Ifo Institute for Economic Research at the University of Munich. https://www.ifo.de/en/publications/2006/working-paper/qualitative-business-surveys-manufacturing-and-industrial.

Bachmann, R., S. Elstner, and E.R. Sims. 2013. "Uncertainty and Economic Activity: Evidence from Business Survey Data". *American Economic Journal: Macroeconomics*, Volume 5, Issue 2: 217-249. https://www.doi.org/10.1257/mac.5.2.217.

Bruno, G., L. Crosilla, and P. Margani. 2019. "Inspecting the Relationship Between Business Confidence and Industrial Production: Evidence on Italian Survey Data". *Journal of Business Cycle Research*,Volume 15, Issue 1: 1-24. https://doi.org/10.1007/s41549-018-00033-4

Bruno, G., and C. Lupi. 2004. "Forecasting industrial production and the early detection of turning points". *Empirical Economics*, Volume 29, Issue 3: 647-671.

Bruno, G., and M. Malgarini. 2002. "An Indicator of Economic Sentiment for the Italian Economy". *ISAE Working Papers*, N. 28/2002. Roma, Italy: Istat.

Cesaroni, T. 2011. "The cyclical behavior of the Italian business survey data". *Empirical Economics*, Volume 41, Issue 3: 747-768. https://doi.org/10.1007/s00181-010-0390-7.

Cesaroni, T., and S. Iezzi. 2017. "The Predictive Content of Business Survey Indicators: Evidence from SIGE". *Journal of Business Cycle Research*, Volume 13, Issue 1: 75-104. https://doi.org/10.1007/s41549-017-0015-8.

Cesaroni, T., L. Maccini, and M. Malgarini. 2011. "Business cycle stylized facts and inventory behaviour: New evidence for the Euro area". *International Journal of Production Economics*, Volume 133, Issue 1: 12-24. https://doi.org/10.1016/j.ijpe.2010.07.044.

Chong, Y.Y., and D.F. Hendry. 1986. "Econometric Evaluation of Linear Macro-Economic Models". *The Review of Economic Studies*, Volume 53, Issue 4: 671-690. https://doi.org/10.2307/2297611.

Claveria, O., E. Pons, and R. Ramos. 2007. "Business and consumer expectations and macroeconomic forecasts". *International Journal of Forecasting*, Volume 23, Issue 1: 47-69. https://doi.org/10.1016/j.ijforecast.2006.04.004.

Fok, D., P.H. Franses, and R. Paap. 2006. "Performance of Seasonal Adjustment Procedures: Simulation and Empirical Results". In Mills, T.C., and K. Patterson (Eds.). *Palgrave Handbook of Econometrics. Volume 1: Econometric Theory*. London, UK: Palgrave Macmillan.

Frale, C., and L. Monteforte. 2011. "FaMIDAS: A Mixed Frequency Factor Model with MIDAS structure". *Temi di Discussione (Working Papers),* N. 788. Roma, Italy: Bank of Italy. https://www.bancaditalia.it/pubblicazioni/temi-discussione/2011/2011-0788/en_tema_788.pdf?language_id=1.

Girardi, A., C. Gayer, and A. Reuter. 2016. "The Role of Survey Data in Nowcasting Euro Area GDP Growth". *Journal of Forecasting*, Volume 35, Issue 5: 400-418. https://doi.org/10.1002/for.2383.

Hansson, J., P. Jansson, and M. Löf. 2003. "Business Survey Data: Do They Help in Forecasting the Macro Economy?". *Sveriges Riksbank Working Paper Series*, N. 151. Stockholm, Sweden: Sveriges Riksbank. https://www.econstor.eu/bitstream/10419/82449/1/wp_151.pdf.

Harvey, D.I., S.J. Leybourne, and P. Newbold. 1998. "Tests for Forecast Encompassing". *Journal of Business & Economic Statistics*, Volume 16, Issue 2: 254-259. https://doi.org/10.2307/1392581.

Krolzig, H.-M., and D.F. Hendry. 2001. "Computer automation of general-to-specific model selection procedures". J*ournal of Economic Dynamics and Control*, Volume 25, Issue 6-7: 831-866. https://doi.org/10.1016/S0165-1889(00)00058-0.

Lemmens, A., C. Croux, and M.G. Dekimpe. 2005. "On the predictive content of production surv33eys: A pan-European study". *International Journal of Forecasting*, Volume 21, Issue 2: 363-375. https://doi.org/10.1016/j.ijforecast.2004.10.004.

Mazzi, G.L., and G. Savio. 2005. "The Seasonal Adjustment of Short time series". *Working Papers and Studies.* Luxembourg: Office for Official Publications of the European Communities. https://ec.europa.eu/eurostat/documents/3888793/5832357/KS-DT-05-002-EN.PDF/54bdfe54-f3b3-49cb-a0e5-9edcc64199ff.

Proietti, T. 2012. "Seasonality, Forecast Extensions and Business Cycle Uncertainty". *Journal of Economic Surveys*, Volume 26, Issue 4: 555-569. https://doi.org/10.1111/j.1467-6419.2010.00660.x.

Van Nieuwerburgh, S., and L. Veldkamp. 2006. "Learning asymmetries in real business cycles". *Journal of Monetary Economics*, Volume 53, Issue 4: 753-772. https://doi.org/10.1016/j.jmoneco.2005.02.003.

An empirical comparison of some outlier detection methods with longitudinal data

Marcello D'Orazio[1]

Abstract

This paper investigates the problem of detecting outliers in longitudinal data. It compares well-known methods in official statistics with some proposals from the data mining or machine learning field, which are based on the distance between observations or binary partitioning trees. The comparison is done by applying them to panel survey data related to different types of statistical units. Traditional methods are relatively simple and enable the direct identification of potential outliers; however, they require specific assumptions. Recent methods provide just a score whose magnitude is directly related to the chance of having an outlier. All the methods require setting a number of tuning parameters; however, the most recent methods show higher flexibility and are sometimes more effective than traditional ones. Additionally, these methods can be applied in the multidimensional case.

Keywords: Boxplot, isolation forest, nearest neighbour distance, outlier detection, panel data.

DOI: https://doi.org/10.1481/ISTATRIVISTASTATISTICAUFFICIALE_1-2-3.2023.06.

1 Marcello D'Orazio (madorazi@istat.it), Italian National Institute of Statistics - Istat.

The views and opinions expressed are those of the author and do not necessarily reflect the official policy or position of the Italian National Institute of Statistics - Istat.
The author would like to thank the anonymous reviewers for their comments and suggestions, which enhanced the quality of this article.

1. Introduction

Data collected in sample surveys as well as data in administrative registers often contain errors that, if not corrected, may affect the accuracy of the final estimates. For these reasons, National Statistical Institutes (NSIs) have always invested considerable resources in verifying the incoming data to detect actual or potential errors. This process is known as *data editing* (or *statistical data editing;* de Waal *et al.*, 2011; sometimes also referred to as *input data validation*) and aims also at identifying *missing values* that are then replaced by *imputed values*[2]; the *imputation* process also permits replacing the values identified as erroneous (deleted and imputed). The editing and imputation sub-process can make use of a variety of statistical methods depending on many factors: the type of data source; the data collection mode; the nature of variables (continuous, categorical, or mixed-type) and the relationship existing between them; the nature of errors and their potential impact on the final estimates, etc.

This paper focusses on the subset of data editing methods tailored to *outlier detection*; "an outlier is a data value that lies in the tail of the statistical distribution of a set of data values" (UNECE, 2000)[3]. The underlying idea is that "outliers in the distribution of uncorrected (raw) data are more likely to be incorrect". For instance, when observing a single continuous variable (household income, firm production, harvested area in a farm, etc.), an outlier can be the outcome of a *measurement error*, *i.e*. the observed value is not the true value (and the true value is not expected to be in the tail of the distribution). An outlier can also be a non-erroneous "extreme" value that, although it has a significant influence on the final estimates, may deserve a "special" treatment in the analysis[4].

This article examines traditional and recent approaches to outlier detection in longitudinal data, where a continuous variable is observed on the same set

2 "Data editing and imputation" is a sub-process in the "Process" phase of the Generic Statistical Business Process Model (GSBPM; https://statswiki.unece.org/display/GSBPM/GSBPM+v5.1).

3 "An outlier is an observation which is not fitted well by a model for the majority of the data. For instance, an outlier may lie in the tail of the statistical distribution or 'far away from the centre' of the data" (Memobust Glossary; https://ec.europa.eu/eurostat/cros/content/glossary_en).

4 In sample surveys, in theory, it is possible to distinguish between a *representative outlier* (*i.e*. a unit in the sample that represents other units in the target population that have similar values) and a *non-representative outlier* (when in the target population no other units are showing similar values; https://ec.europa.eu/eurostat/cros/content/glossary_en).

of units at different time points. In the case of panel surveys, official statistics can observe households, firms, or agricultural holdings. The following Section briefly describes well-known outlier detection methods. It illustrates some nonparametric approaches suggested in the field of data mining or machine learning, which have great potential when applied to this specific setting and, more generally, in official statistics. Section 3 compares the outcomes of the application of the reviewed methods to panel data related to different types of statistical units, often investigated in surveys conducted by NSIs. Finally, Section 4 drafts some concluding remarks and future areas of work.

2. Some outlier detection methods for longitudinal data

When a given quantitative non-negative variable *Y* is observed repeatedly over time on the same set of units, we can expect a high correlation between subsequent measurements; this feature represents useful information in setting up an efficient outlier detection procedure. It becomes even more relevant when the objective is to estimate the change over time of a population parameter related to *Y*.

Formally, let $(y_{t1}, y_{t2}, \dots, y_{tn})$ be the values of *Y* observed at time *t* on a set of *n* units, being $y_{ti} \geq 0$; the ratio $r_i = y_{t_2 i}/y_{t_1 i}$ denotes the "individual change" from time t_1 to time t_2 for unit *i*, being $t_1 > t_2$. Data editing literature suggests various methods to check whether the individual change (r_i) is too large or too low (*e.g.* the theme "Editing for Longitudinal Data" in Eurostat, 2014); a very popular one was suggested by Hidiroglou and Berthelot (1986), and its characteristics are summarised in the following 2.1.

2.1 Hidiroglou-Berthelot method for outlier detection

Hidiroglou and Berthelot (1986) suggested examining the empirical distribution of the ratios $r_i = y_{t_2 i}/y_{t_1 i}$ $(i = 1,2,\dots,m;$ being $m \leq n,$ after discarding 0s and missing values, if any, in both $y_{t_2 i}$ and $y_{t_1 2}$). In particular, they first transform the ratios in the following manner:

$$s_i = \begin{cases} 1 - \frac{r_M}{r_i}, & \text{if} \quad 0 < r_i < r_M \\ \frac{r_i}{r_M} - 1, & \text{if} \qquad r_i \geq r_M \end{cases} \tag{1}$$

such that s_i=0 when a ratio is equal to the median of ratios $(r_i = r_M)$; then, to account for the magnitude of data and give more "importance" to units involving high values of the *Y* variable, suggest to derive the following *score*:

$$E_i = s_i\left[\max\left(y_{t_1 i}, y_{t_2 i}\right)\right]^U \tag{2}$$

where *U* can range from 0 to 1 $(0 \leq U \leq 1)$ and controls the role of the magnitude in determining the importance associated with the centred ratios, a common choice consists of setting *U* = 0.5.

Identification of potential outliers relies on the assumption that the scores are approximately distributed according to a Gaussian distribution; in practice, the parameters of the Gaussian distribution are estimated using robust methods and units outside the interval:

$$[E_M - C \times d_{Q1}, E_M + C \times d_{Q3}] \tag{3}$$

are identified as potential outliers. In expression (3):

$$d_{Q1} = max\left(E_M - E_{Q1}, |A \times E_M|\right) \qquad d_{Q3} = max\left(E_{Q3} - E_M, |A \times E_M|\right) \tag{4}$$

being $\boldsymbol{E_{Q1}}$, $\boldsymbol{E_M}$, $\boldsymbol{E_{Q3}}$ and the quartiles of the *E* scores. The constant *A* is a positive small quantity (suggested $A = 0.05$) introduced to overcome cases of $\boldsymbol{E_M = E_{Q3}}$ or $\boldsymbol{E_M = E_{Q1}}$ that may occur when the ratios are too concentrated around their median. The parameter *C* determines how far from the median the bounds should be; commonly suggested values are $C = 4$ or $C = 7$, but larger values can be considered, depending on the tails of the distribution of the *E* scores. In practice, the bounds (3) allow for a slight skewness in the distribution of the *E* scores.

Recently, Hidiroglou and Emond (2018) suggested replacing $\boldsymbol{E_{Q1}}$ and $\boldsymbol{E_{Q3}}$ with, respectively, the percentiles $\boldsymbol{E_{P10}}$ and $\boldsymbol{E_{P90}}$ when a large proportion of units (>1/4) share the same value of the ratio, since in this case the "standard" method would detect too many observations as potential outliers.

Practically, the decision about the values of the "tuning" constants *U* and *C* is not straightforward and requires a graphical investigation of the distribution of scores, as well as different attempts with alternative values of both the constants. A helpful practical suggestion is to start inspecting the (suspicious) ratios by sorting them in decreasing order with respect to the absolute value of the score ($|E_i|$). Hidiroglou and Emond (2018) also suggested an additional graphical inspection procedure.

In the R environment (R Core Team, 2022), the Hidiroglou Berthelot (HB) procedure is implemented by the function `HBmethod()` available in the package *univOutl* (D'Orazio, 2022), which also includes graphical facilities for inspecting the scores, in line with Hidiroglou and Emond's recommendations (2018). In addition, *univOutl* has facilities to identify outliers in univariate cases with methods based on robust location and scale estimates of the parameters of the Gaussian distribution.

2.2 Nonparametric methods

The nonparametric methods for outlier detection are very popular because they do not introduce an explicit assumption on the underlying distribution. For the sake of simplicity, when describing these methods, it is assumed that the problem of detecting outliers arises when a generic continuous variable *X* is observed on a set of *m* observations. The following Subsections describe some well-known approaches using boxplots, as well as recent proposals in the field of data mining and machine learning, particularly *distance and tree-based methods*.

2.2.1 Outlier detection with boxplots

Drawing a *boxplot* (*box-and-whisker* plot) is a popular approach to outlier detection:

$$[Q1_x - c \times IQR_x; \;\; Q3_x + c \times IQR_x] \tag{5}$$

where IQR_x is the inter-quartile range $(IQR_x = Q3_x - Q1_x)$ and, usually, $c = 1.5, \; 2 \text{ or } 3$; units outside the bounds (*whiskers*) are considered outliers.

To account for moderate skewness, Hubert and Vandervieren (2008) suggested an "adjusted" boxplot:

$$[Q1_x - 1.5 \times exp(aM) \times IQR_x; \;\; Q3_x + 1.5 \times exp(bM) \times IQR_x] \tag{6}$$

being *M* the *medcouple* measure of skewness $(-1 \leq M \leq 1$; Brys *et al*., 2004) that when greater than 0 indicates positive skewness and requires setting *a* = -4 and *b* = 3 in expression (6); on the contrary, with negative skewness (*M* < 0) it is suggested to set a = and b = 4. The authors claim that the adjusted boxplot fences in (6) work with moderate skewness, *i.e*. $-0.6 \leq M \leq 0.6$. Unfortunately, the adjusted boxplot permits only to set *c* = 1.5, and it is not possible to use alternative values.

D'Orazio (2022) implemented standard and adjusted boxplots in the function `boxB()` of the R package *univOutl*. Additionally, the function `HBmethod()` enables the application of the adjusted boxplot to the *E* scores.

2.2.2 Distance-based outlier detection

The idea of using distance measures in outlier detection is a direct consequence of the fact that we search for observations that are far from the centre of the data. In practice, distance-based outlier detection methods search for an observation that has very few other observations close to it; the fewer observations close to a unit, the higher the chance that it is an outlier. Knorr and Ng (1998) suggest identifying an outlier as an observation that has fewer than *k* observations at a distance less than or equal to a threshold δ. This approach requires deciding: (i) how to measure the distance, (ii) the distance threshold δ, and (iii) the *k* parameter. The first two choices are strictly related and relatively simple in the univariate setting (but not in the multidimensional case), as different distance functions may leave the set of nearest neighbours of a given unit unchanged.

Knorr and Ng's approach (1998) does not provide a ranking for the potential outliers. To overcome this difficulty, Ramaswamy *et al*. (2000) suggest identifying the potential outliers by calculating the *k* nearest neighbour (*k*-NN) distance; in practice, if $\boldsymbol{d_i^{(k)}}$is the distance of the ith from its *k*-nearest neighbour, the units showing the largest values of $\boldsymbol{d_i^{(k)}}$ are potential outliers. This simple approach can be computationally demanding in the presence of many observations and variables; however, some algorithms simplify the search (Hautamäki *et al*., 2004). The problem becomes much simpler in the univariate case, where the initial ordering of the units reduces the computational effort.

The *k*-NN distance proposed by Ramaswamy *et al*. (2000) is a very popular approach, and many variants are available. A well-known extension assigns to each unit a "weight" consisting of the sum of its distance from the corresponding *k* nearest neighbour observations (Angiulli and Pizzuti, 2002):

$$\boldsymbol{\omega_i^{(k)} = \sum_{j=1}^{k} d_i^{(j)}} \tag{7}$$

Hautamäki *et al.* (2004) suggest using the average. Common choices for the parameter *k* are 5 or 10; however, the literature does not provide a rule of thumb. Campos *et al.* (2016) note that the sum of distances makes the scores less sensitive to the value of *k*. Obviously, if *k* is too large, then the weight may become quite large even for non-outlying observations, since, as shown by expression (7), the final score will be the result of a sum of a larger number of terms.

In general, distance-based outlier detection methods are strictly related to outliers' detection based on kernel density estimation techniques; this paper will not address such methods, but it is worth noting that when *k*-NN is applied to density estimation problems, a possible rule of thumb consists in setting $k = ceiling(\sqrt{m})$ and, more in general, $k \sim m^{4/5}$.

The main drawback with *k*-NN methods is that they do not directly identify the potential outliers, like boxplots or the HB method; instead, they provide a summary score (distance or "weight") whose magnitude indicates the chance of being an outlier; the larger the score, the higher is the chance that a given observation is an outlier. To identify a possible threshold such that observations with a score above the threshold are identified as potential outliers, Hautamäki *et al*. (2004) suggest:

$$u_0 = \varepsilon \times max\left[u_{(i+1)}^{(k)} - u_{(i)}^{(k)}\right],\ i = 1,2,\dots,m-1 \tag{8}$$

where $u_i^{(k)} = \omega_i^{(k)}$ or $u_i^{(k)} = d_i^{(k)}$, $u_{(i+1)}^{(k)} \geq u_{(i)}^{(k)}$ and ε is a user-defined constant $0 < \varepsilon < 1$. This rule, introduces an additional parameter to set (ε); practically, a graphical inspection of the ordered scores $u_i^{(k)}$ can be more effective: once sorted them increasingly, good candidate thresholds the values corresponding to "jumps" in the plot (abnormal increase in the score).

The approach by Knorr and Ng (1998) is closely related to the DBSCAN (*Density-based spatial clustering applications with noise*) clustering algorithm (Ester *et al*., 1996), where the observations not "reachable" by any other observation are identified as *noisy* observations (outliers). The "reachability" depends on a distance threshold δ; in practice, two observations i and j are *directly reachable* if their distance is less than, or equal to, δ $(d_{i,j} \leq \delta)$. At the same time, they are only *reachable* if there is a path of three or more observations to go from i to j, where each couple of units in the path is directly reachable. The DBSCAN algorithm requires setting also a value for g that is needed to identify the *core observations*, *i.e*. observations that have at least $k = g - 1$ distinct units at a distance smaller than or equal to δ. To identify a value for δ, it is suggested to plot the *k*-NN distances in increasing order and set δ equal to the distance where the plot shows a jump.

In R, some distance-based methods for outlier detection are implemented in the package *DDoutlier* (Madsen, 2018), although *k*-NN distance is calculated in many other R packages; the package *dbscan* (Hahsler *et al*., 2019 and

2022) implements the DBSCAN clustering algorithm but has also facilities to calculate the *k*-NN distance efficiently.

2.2.3 Outlier detection with isolation forest

The *isolation forest* is an unsupervised decision-tree-based algorithm that consists of fitting an ensemble of *isolation trees* (Liu *et al*., 2008 and 2012). The underlying idea is that outlying observations have a higher chance of being separated from the others in one branch of the partitioning tree, with relatively few splits. In the univariate case an arbitrary threshold $\boldsymbol{x_o}$ is selected at random within the range of X $([min(x_i), max(x_i)])$ and all the observations are divided into two groups according to whether they show higher or lower values than $\boldsymbol{x_o}$. This randomised splitting process is applied recursively (*i.e.* divide the units into two groups then repeat the process in each group, and so on) until no further split is possible or until meeting some other criteria. The final outcome is an isolation tree where the more observations show similar *X* values, the longer (more splits) it will take to separate them into small groups (or alone) compared to less occurring *X* values; for this reason, the *isolation depth* (number of splits needed to isolate a unit) can be considered as a tool for detecting outliers.

Since a high variability would characterise the isolation depth estimated in a single isolation tree, its reduction can be achieved by building an ensemble of isolation trees – the isolation forest – and then deriving the final score by averaging over the fitted trees. As in random forests, each single isolation tree is fit on a bootstrap sample of q $(q < m)$ observations randomly selected. In the Liu *et al*. proposal (2008 and 2012), the partitioning stops when a node has only one observation, or all units in a node have the same values (in some cases, it is also introduced a maximum value for the tree height, *e.g.* $l_{max} = ceiling(log_2(q))$).

Formally, if $h(x_i)$ is the *path-length or depths*, *i.e*. the number of splits to reach the *i*-th observation in a fitted tree, Liu *et al.* (2008, 2012) suggest to associate to each observation the following score:

$$u_i = 2^{-\frac{E[h(x_i)]}{c(q)}} \tag{9}$$

where $E[h(x_i)]$ is the average path length across the ensemble of the fitted trees and

$$c(q) = 2 \times H(q-1) - 2\,\frac{q-1}{q} \qquad (10)$$

being $H(\cdot)$ the harmonic number. The Authors demonstrate that the resulting score ranges from 0 to 1 ($0 < u_i \leq 1$) being a monotonic function of $h(x_i)$ and, in particular

- when $E[h(x_i)] \rightarrow q - 1$ then $u_i \rightarrow 0$;
- when $E[h(x_i)] \rightarrow c(q)$ then $u_i \rightarrow 0.5$;
- when $E[h(x_i)] \rightarrow 0$ then $u_i \rightarrow 1$.

Practically, scores close to 1 indicate observations with a very short average path length that tend to be isolated earlier than the other ones and therefore denote outlying observations. As a consequence, setting a threshold score u_0 will return as outliers all the units having a score $u_i > u_0$. Generally, it is suggested to consider $u_0 > 0.5$, but a graphical inspection of the ordered scores can be beneficial in deciding u_0.

The isolation forest is very efficient and can handle multi-modal distributions. It requires setting two tuning parameters, the subsample size q and the number of trees to fit. In the first case, Liu *et al*. (2008, 2012) claim that even a small subsample size ($q = 256$) can work with massive datasets, while at least 100 trees should be fitted; this latter number should be increased when the achieved scores are on average quite below 0.5, as this may point out a problem of unreliable estimation of the average path length. It is worth noting that the standard method proposed by Liu *et al*. (2008, 2012) is also developed to handle outlier detection in a multidimensional framework, where the creation of each tree requires a recursive random selection of one of the available variables and its corresponding random splitting point. In the univariate case, with a single variable, there is just the random selection of the splitting point, and consequently, there is no need to grow a large number of trees. It is worth noting that, in the multivariate case, the standard algorithm (Liu *et al.*, 2008, 2012) essentially consists of an ensemble of results related to the application of the isolation forest independently to each variable. To compensate for this drawback, Hariri *et al*. (2021) proposed an *extended isolation forest* that, in the branching step, considers jointly two or more variables; for instance, when two variables are randomly selected, then the algorithm partitions repeatedly the units according to a regression line whose intercept and slope are randomly generated each time.

In R, the standard isolation forest is implemented in the *solitude* package (Srikanth, 2021), while the *isotree* package (Cortes, 2022) implements the "base" isolation forest algorithm, as well as some of its variants.

3. Application of the chosen methods to some data from panel surveys

This section examines the performance of the methods presented in the previous Section when applied to various datasets related to panel or pseudo-panel surveys, as described in Table 3.1.

Table 3.1 – Datasets used in the experiments

Dataset/survey	Number of units	Type of units	Description
RDPerfComp	509	Firms	R&D performing US manufacturing; yearly observations from 1982 to 1989 of the following variables: production, labour and capital (a).
RiceFarms	171	Farms	The Indonesian rice farm dataset comprises 171 farms that produce rice, which were observed six times. Several variables are available, including hectares of cultivated area, gross output of rice in kilograms, and net output, among others (b).
Wages	595	Individuals	A panel of 595 individuals from 1976 to 1982, taken from the Panel Study of Income Dynamics (PSID); many variables available (see footnote 2)
Survey on Household Income and Wealth (SHIW)	3,804	Households	Subset of panel households observed in 2014 and 2016; many variables available: net income, consumption, wealth, etc. (c).

Source: Author's processing
(a) https://www.nuffield.ox.ac.uk/users/bond/index.html. See also the R package pder https://CRAN.R-project.org/package=pder.
(b) R package plm https://cran.r-project.org/package=plm.
(c) Bank of Italy, Survey on Household Income and Wealth, years 2014 and 2016. Public use anonymised microdata distributed for research purposes https://www.bancaditalia.it/statistiche/tematiche/indagini-famiglie-imprese/bilanci-famiglie/distribuzione-microdati/index.html?com.dotmarketing.htmlpage.language=1.

In practice, in each dataset, the HB procedure is applied to the chosen variable and the resulting $\boldsymbol{E_i}$ scores, calculated using expression (2), become the input of the outlier detection techniques listed in the first column of Table 3.2, whose corresponding parameters/tuning constants are given in the second column of the table. All analyses were carried out in the R environment. Columns 3 and 4 in Table 3.2 provide details related to the chosen R package, function, and corresponding arguments (arguments not explicitly mentioned are set equal to their default values)[5].

5 The used R code can be found in the GitHub repository: https://github.com/marcellodo/univOutl.

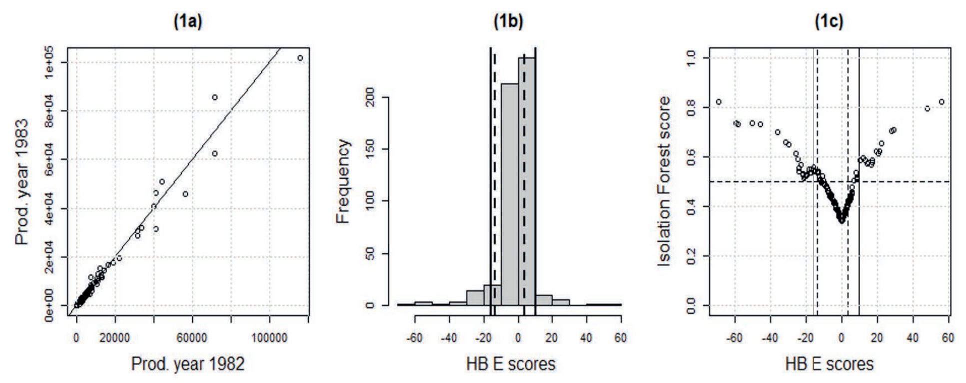
Table 3.2 – Methods, R functions and corresponding tuning parameters

Method	Parameters	R function and package	Arguments in the R function
Hidiroglou-Berthelot ("HB")	Quartiles (deciles for Wages dataset); $U = 0.5$; $C = 7$; $A = 0.05$	`HBmethod()` in *univOutl*	`U=0.5, A=0.05, C=7, pct=0.25 (pct=0.10` for Wages dataset)
Skewness-adjusted boxplot ("SABP") (see eqn. (6))		`boxB()` in *univOutl*	`k=1.5, method='adjbox'`
Isolation Forest ("IF")	No subsampling; 500 trees	`isolation.forest()` in isotree	`ntrees=500`
DBSCAN	Three runs with different values of g (6, 11,16) and different thresholds for δ (decided after graphical inspection of the sorted (g-1)-NN distances calculated on the $\boldsymbol{E_i}$)	`dbscan()` in *dbscan*	`minPts=6, minPts=11, minPts=16` `eps` set equal to the decided for each combination of `minPts` and the various datasets
k-NN outlier detection ("k-NN-dist")	Three runs with different values of k (5, 10,15)	`kNNdist()` in *dbscan*	`k=5, k=10, k=15`
k-NN weights ("k-NN-weight"; see expression (7))	Three runs with different values of k (5, 10,15)	`kNNdist()` in *dbscan*	`k=5, k=10, k=15, all=TRUE`

Source: Author's processing

The variable examined in the RDPerfComp dataset is the firms' production in 1983 compared to 1982. Figure 3.1 reports the observed scatterplot (1a); plot (1b) shows the distribution of the HB scores ($\boldsymbol{E_i}$) and the vertical continuous lines indicate the HB bounds provided by (3). In contrast, the dashed lines represent the fences of the skewness-adjusted boxplot, as provided by equation (6). The histogram shows a moderate negative skewness ($M = -0.2338$), and the SABP fences identify a higher number of potential outliers if compared to HB (with $C = 7$ and $A = 0.5$).

Figure 3.1 - Scatterplot of the data related to firms' production (1a), distribution of the HB scores (1b), and relation between HB and IF scores (1c)

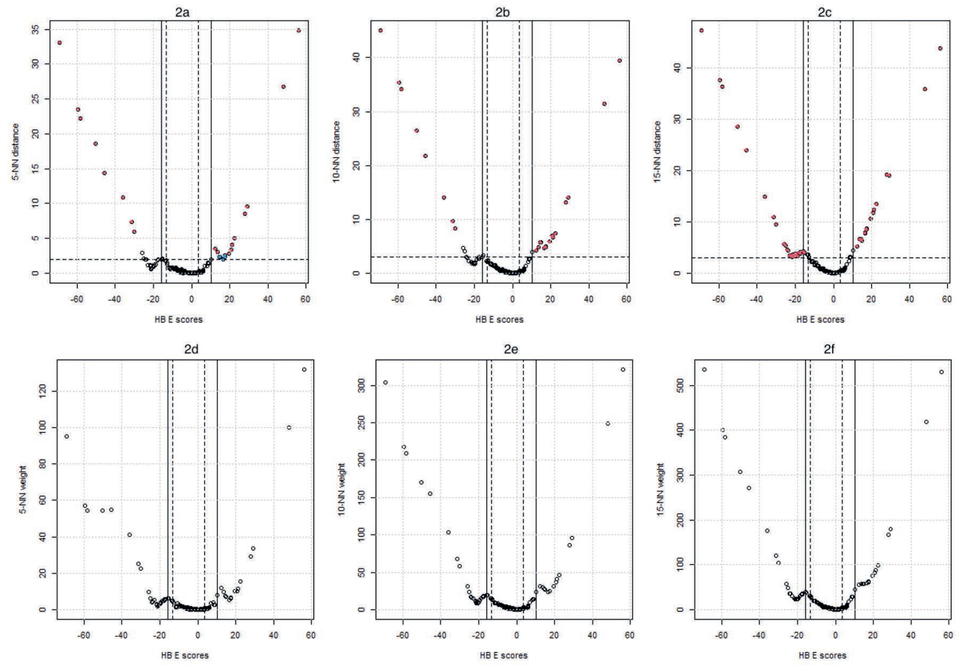


Source: Author's processing

The right-side plot (1c) reports the scatterplot of the IF scores (u_i) vs.E_i. Many observations identified as outliers by the HB method have an IF score slightly above 0.5, while the observations with the maximum observed IF score (slightly beyond 0.8, with a maximum achievable score of 1) are relatively few.

Figure 3.2 reports the scatterplots of scores provided by k-NN-dist (2a-2c) and k-NN-weight (2d-2f) compared to the input HB scores (E_i). Scatterplots (2a-2c) also show the findings of DBSCAN with respectively $\delta = 2$, $\delta = 3$ and $\delta = 3$. In these plots, the red-colour points indicate the noisy points (outliers), while the blue-colour ones form a separate cluster of observations, far from most of the observations that, however, are not identified as outliers. In this example, the outliers returned by DBSCAN are always fewer than those provided by the standard HB method. More generally, 5-NN and 10-NN are more effective than 15-NN distance in identifying potential outliers (units with higher distance).

Figure 3.2 - Firms' production data, relationship between HB and scores provided by the *k*-NN methods



Source: Author's processing

Scatterplot (2d), (2e), and (2f) compare the E_i with the sum of the *k*-NN distances (*k*-NN-weight). If compared to the "standard" *k*-NN-dist, the sum of the distances (weight), as expected, seems less sensitive to the value of *k* (Campos *et al*., 2016) and helps more in detecting the potential outliers (units with the highest weight $\omega_i^{(k)}$), in particular when $k = 5$ and $k = 10$

Table 3.3 shows the estimated Kendall's *tau* correlation coefficient between the various scores obtained at the end of the different procedures for outlier detection (for HB, it is considered the absolute value $|E_i|$). Correlations are relatively high, indicating a high concordance between rankings of the scores produced by the different methods. IF scores are highly correlated with $|E_i|$. Concordance between $|E_i|$ and the scores provided by the application of *k*-NN methods increases with increasing values of *k*.

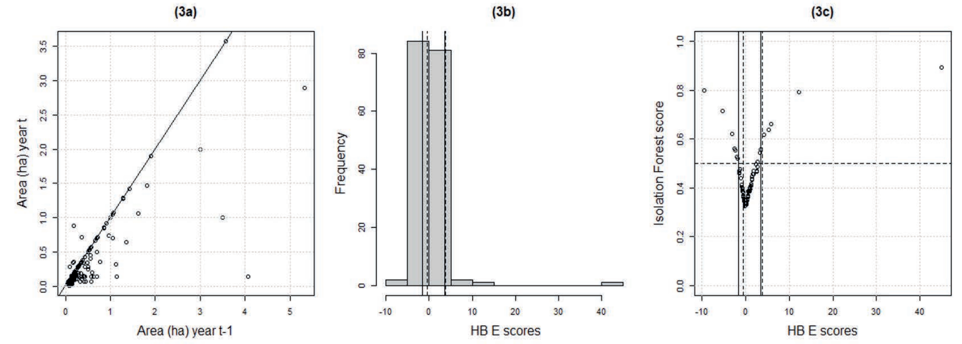
Table 3.3 – Kendall's correlation between the scores assigned to the firms

	IF	5-NN-dist	10-NN-dist	15-NN-dist	5-NN-weight	10-NN-weight	15-NN-weight
|E|	0.8984	0.7672	0.8269	0.8689	0.7435	0.8094	0.8471
IF		0.8091	0.8665	0.9009	0.7845	0.8544	0.8920
5-NN-dist			0.829	0.7974	0.9045	0.8931	0.8587
10-NN-dist				0.8913	0.8050	0.9198	0.9414
15-NN-dist					0.7717	0.8596	0.9144
5-NN-weight						0.8712	0.8296
10-NN-weight							0.9400

Source: Author's processing

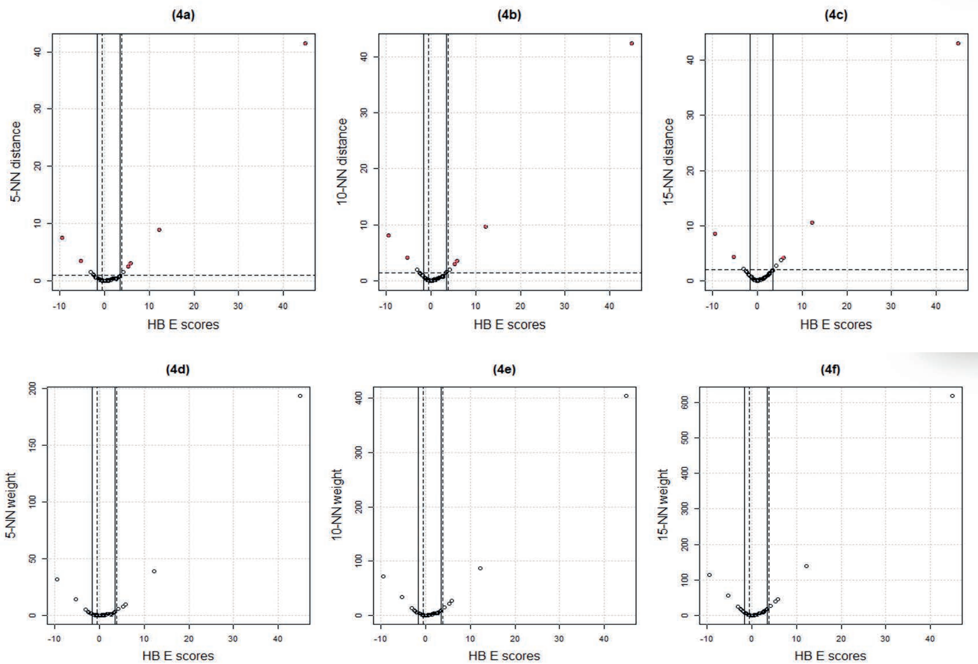
Figures 3.3 and 3.4 summarise the results of the different outlier detection procedures when applied to the area harvested for rice production in the observed farms (5th time occasion vs. the 4th) listed in the RiceFarm dataset. Histogram (3b) indicates a moderate positive skewness (M=0.3697). The fences of the SABP are close to the bounds of HB intervals, particularly on the right tail of the distribution; as expected, SABP appears to better account for moderate positive skewness. The identified outlying farms are relatively few and, in general, show an IF score greater than 0.6 (with few exceptions, located in the left tail). DBSCAN with the chosen distance thresholds ($\delta = 1$, $\delta = 1.5$ and $\delta = 2$, respectively; decided by after graphical inspection of the sorted *k*-NN distances) identifies quite a few outliers, slightly less than those identified by HB or SABP. Plots related to *k*-NN methods (4a-4c and 4d-4f) show that there are a few outlying farms with scores (*k*-NN distance or *k*-NN weight) that are not close to those of the majority of farms.

Figure 3.3 - Scatterplot of the area for rice production (3a), distribution of the HB scores (3b), and relation between HB and IF scores (3c)



Source: Author's processing

Figure 3.4 - Rice-growing area in farms, relationship between HB and scores provided by the k-NN methods



Source: Author's processing

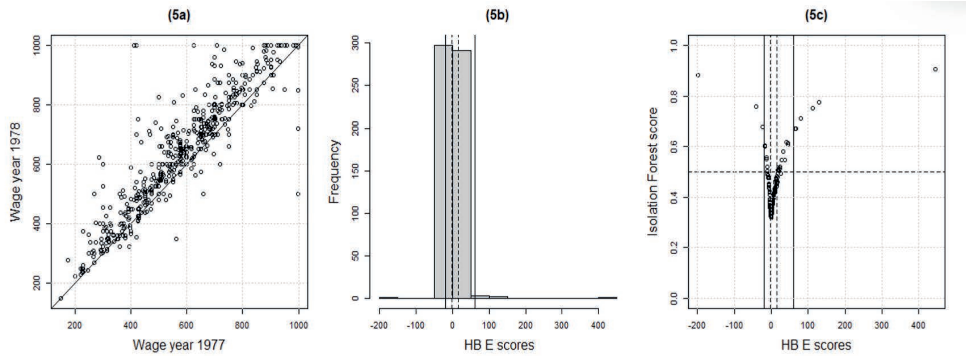
Table 3.4 shows that, also in this case, the IF score is the one with higher correlation (measured in terms of Kendall's tau) with the absolute value of the HB scores ($|E_i|$). Rankings based on IF scores tend to agree more with those provided by 10-NN and 15-NN methods. In general, correlations are all relatively high.

Table 3.4 - Kendall's correlation between the scores assigned to the farms producing rice

	IF	5-NN-dist	10-NN-dist	15-NN-dist	5-NN-weight	10-NN-weight	15-NN-weight
|E|	0.8627	0.7320	0.8285	0.8306	0.7064	0.8205	0.8401
IF		0.8084	0.9080	0.9125	0.7798	0.9054	0.9299
5-NN-dist			0.8071	0.7934	0.8978	0.8775	0.8306
10-NN-dist				0.9154	0.7624	0.9105	0.9480
15-NN-dist					0.7595	0.8890	0.9449
5-NN-weight						0.8333	0.7910
10-NN-weight							0.9392

Source: Author's processing

Figure 3.5 - Scatterplot of the individuals' wages (5a), distribution of the HB scores (5b), and relation between HB and IF scores (5c)

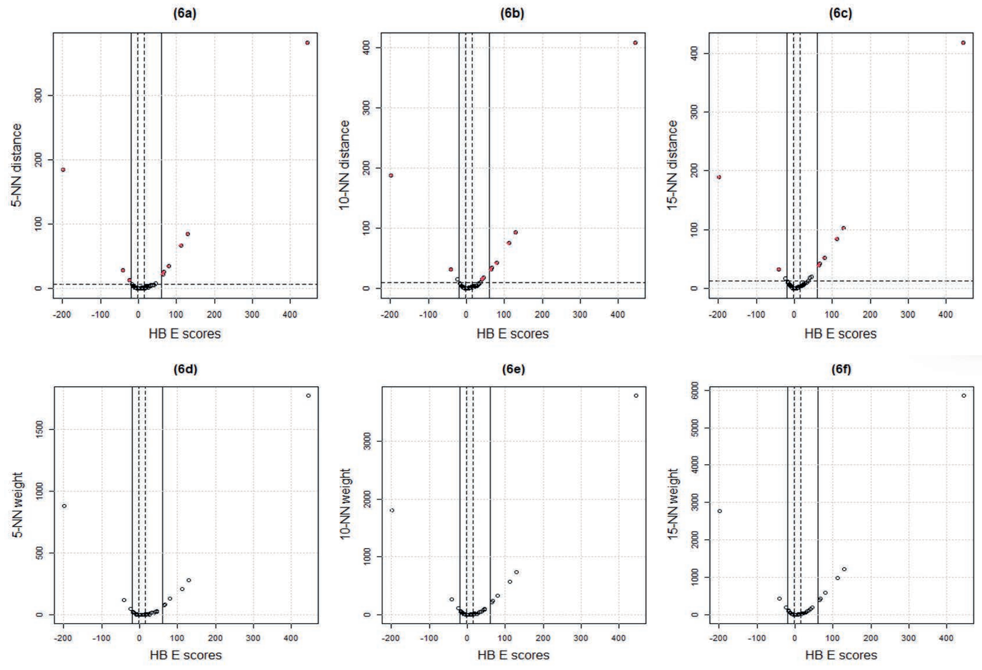


Source: Author's processing

Figures 3.5 and 3.6 show the results obtained by applying the investigated outlier detection methods when analysing the change in individuals' wages from 1977 to 1978, reported in the Wages dataset. Plot (5a) shows that there is an increase in wage for a large subset of individuals. The distribution of the HB E_i scores is positively skewed (M=0.3162), leading to identification of relatively few outliers; in this case, since there is a high concentration of the E_i around the median, in expression (4) it was decided to replace E_{Q1} and

$\boldsymbol{E_{Q3}}$ with respectively $\boldsymbol{E_{P10}}$ and $\boldsymbol{E_{P90}}$, as suggested by Hidiroglou and Emond (2018). This is the reason for the large discrepancy between the HB bounds and those provided by the SABP.

Figure 3.6 - Wages data, relationship between HB and scores provided by the k-NN methods



Source: Author's processing

Outliers identified by the HB method are individuals with an IF score of 0.7 or greater. DBSCAN returns the same outliers identified by HB, except g = 6 (k = 5) (scatterplot 6b) where some additional individuals are identified as outliers. In general, scores provided by the methods based on k-NN show clearly identifiable potential outliers that generally correspond to those identified by the HB procedure.

On average, the estimated correlations reported in Table 3.5 are lower than those calculated with other datasets, indicating that in this case, the rankings provided by the scores do not fully agree. As in other cases, the IF scores are those with higher correlation with the starting $|\boldsymbol{E_i}|$.

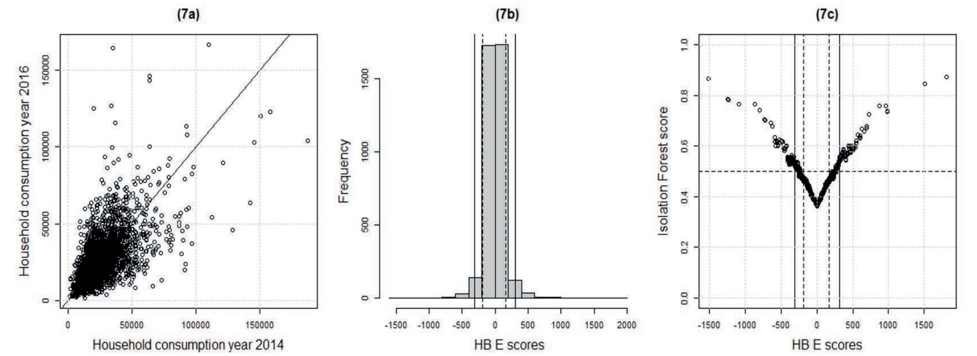
Table 3.5 – Kendall's correlation between the scores assigned to the individual wages

	IF	5-NN-dist	10-NN-dist	15-NN-dist	5-NN-weight	10-NN-weight	15-NN-weight
|*E*|	0.8101	0.5775	0.7045	0.7698	0.5420	0.6654	0.7254
IF		0.6376	0.7695	0.8371	0.6068	0.7426	0.8061
5-NN-dist			0.7287	0.6678	0.8617	0.8377	0.7629
10-NN-dist				0.8396	0.6712	0.8727	0.9200
15-NN-dist					0.6225	0.7861	0.8803
5-NN-weight						0.7836	0.7088
10-NN-weight							0.8976

Source: Author's processing

Figure 3.7 and 3.8 summarise the analyses done on household consumption observed in 2014 and 2016 for the panel component of the Survey on Household Income and Wealth (SHIW) carried out by the Bank of Italy (the survey is biennial). The histogram shows $\boldsymbol{E_i}$ having a symmetric distribution ($M = -0.024$) and the HB bounds ($C = 7$ and $A = 0.5$) determine a relatively high number of outlying households. Similarly, the number of households with IF scores greater than 0.6 is non-negligible, but this subset reduces significantly if the threshold is set to $\boldsymbol{u_0} = 0.7$.

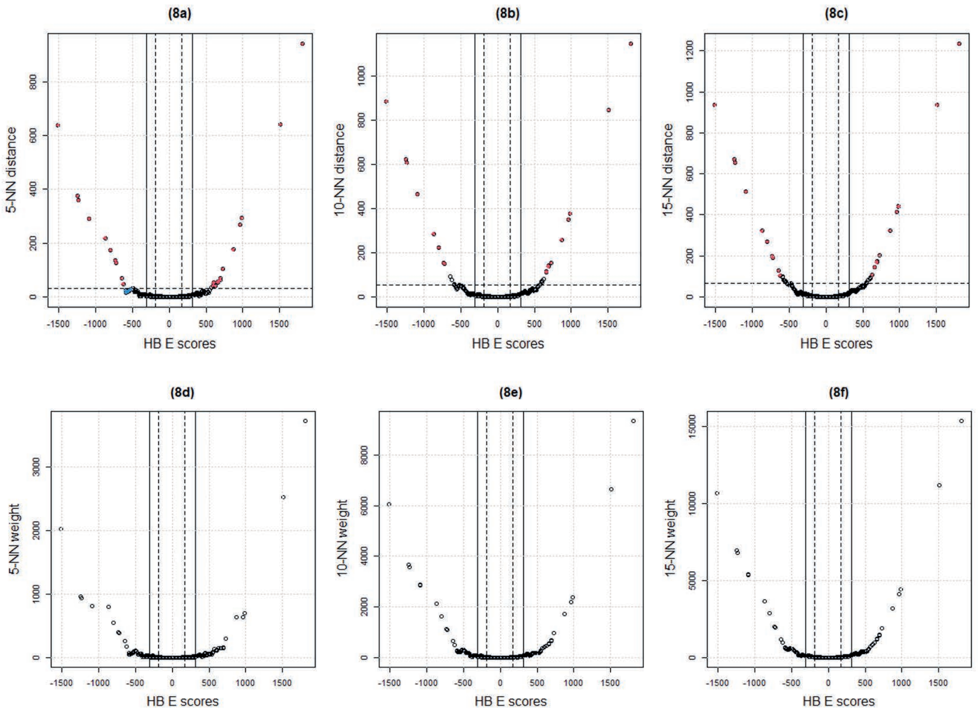
Figure 3.7 - Scatterplot of the household consumption (7a), distribution of the HB scores (7b), and relation between HB and IF scores (7c)



Source: Author's processing

DCSCAN clustering procedure (plots 8a-8c; with δ = 30, δ = 55 and δ = 65, respectively; decided by after graphical inspection of the sorted *k*-NN distances) returns a number of outliers much smaller if compared to the HB method. More generally, all the plots related to *k*-NN distances or *k*-NN weight (8d, 8f) return final scores that facilitate the identification of the outlying observations.

FFigure 3.8 - Household consumption data, relationship between HB and scores provided by the k-NN methods



Source: Author's processing

Kendall's correlations in Table 3.6 show the same tendency highlighted in other situations.

Table 3.6 - Kendall's correlation between the scores assigned to the individual household consumption

	IF	5-NN-dist	10-NN-dist	15-NN-dist	5-NN-weight	10-NN-weight	15-NN-weight
|*E*|	0.9340	0.6702	0.7519	0.7896	0.6493	0.7268	0.7668
IF		0.7048	0.7877	0.8232	0.6833	0.7663	0.8083
5-NN-dist			0.7751	0.7417	0.8737	0.8677	0.8158
10-NN-dist				0.8610	0.7390	0.8820	0.9180
15-NN-dist					0.7129	0.8204	0.8965
5-NN-weight						0.8406	0.7845
10-NN-weight							0.9169

Source: Author's processing

It is worth noting that all scatterplots comparing the HB and IF scores show a "V" shaped diagram except Figure 3.1 (1c) (firms' production), where the E scores show an asymmetric distribution with moderate negative skewness ($M = -0.2338$) but quite "long" tails. The rule of thumb, which identifies units with an IF score greater than 0.5 in SHIW and firms' datasets as potential outliers, returns a relatively high fraction of potential outliers compared to others. This outcome suggests that such a rule should be applied carefully, rather than automatically.

When comparing the HB E scores with those provided by k-NN and "k-NN weight", the scatterplots show a kind of "U" shaped curve with some irregularities depending on the asymmetry in the distribution of the E scores; an exception is again demonstrated by the firms' production data (Figure 3.2). In general, all these scatterplots exhibit some differences when passing from $k = 5$ to $k = 10$. In comparison, shapes remain almost the same for $k = 10$ and to $k = 15$ (obviously, the magnitude of the distance-based scores increases by increasing the values of k), indicating that increasing the value of k too much may not be helpful. DBSCAN is closely related to k-NN since g = $k + 1$, and the analysis of the k-NN distances is required to identify a threshold (parameter δ); it is not a simple task and we opted for a subjective choice guided by a graphical inspection instead of using expression (8) which would require setting the additional tuning constant ε; it is worth noting that for each of the considered datasets the obtained results remain almost stable when varying the combination of the tuning parameters (g and δ); more in general, it seems that this approach returns a relatively small number of observations having however a high chance of being outliers.

4 Conclusions

This paper compares traditional and recent approaches to detecting outliers with longitudinal data, a relatively simple situation that can be practically addressed by applying univariate outlier detection methods. The traditional approaches considered in this study, the HB method and the boxplot, are also popular in official statistics because they have the advantage of permitting a direct identification of the potential outliers (units outside the estimated bounds). The HB method requires setting a series of tuning parameters depending also on the observed distribution of the scores ($\boldsymbol{E_i}$) derived by transforming the initial ratios ($r_i = y_{t_2 i}/y_{t_1 i}$); the method assumes an approximate Gaussian distribution for $\boldsymbol{E_i}$, allowing for slight skewness, but choosing the tuning parameters (to derive the $\boldsymbol{E_i}$ and the final bounds) may require more attempts. Skewness-adjusted boxplot does not explicitly assume a distribution for $\boldsymbol{E_i}$ (apart from that of working with a unimodal unknown distribution) and allows for a moderate skewness; on the contrary, it is not flexible enough as the bounds become too narrow with empirical distributions showing very long tails.

In the wide range of nonparametric methods for outlier detection developed in the fields of data mining and machine learning, we believe that those based on k-NN distances and isolation forests can be efficient and able to handle panel survey data collected in NSIs. These methods offer more flexibility than traditional ones, as they can adapt to different empirical distributions. They ultimately assign a score to each observation, where the larger the score, the higher the chance of being an outlier. This is also their major drawback because it's up to the practitioner to set a threshold such that units with a score beyond it are identified as potential outliers. Only DBSCAN ends up with a clear identification of outliers, but the price to pay is that of setting a threshold for the distance, in addition to the value of *g*. In the case studies considered in this comparison, with the chosen combination of input parameters, this approach generally returns a smaller number of potential outliers compared to traditional techniques and k-NN. For these reasons, DBSCAN seems preferable to k-NN methods; also because it permits the capture of "non-standard" distributions of the $\boldsymbol{E_i}$.

Setting the starting tuning parameters is simpler in the case of the isolation forest, where the practitioner should decide on the size of the bootstrap sample and the number of trees to grow, guided by some rule of thumb mentioned in the literature. The isolation forest has the additional advantage of producing scores ranging in the [0,1] interval, whose midpoint (0.5) represents a good initial candidate for setting a threshold.

In general, the great advantage of “new” nonparametric methods is that they are designed to work also in the multidimensional setting, in contrast to the HB and the boxplot. This is an appealing feature in official statistics, where the data sources often include many variables collected on the same set of units. Additional investigation is, however, required to better understand the pros and cons of these relatively “new” nonparametric methods.

References

Angiulli, F., and C. Pizzuti. 2002. "Fast Outlier Detection in High Dimensional Spaces". In Elomaa, T., H. Mannila, and H. Toivonen (Eds.). *Principles of Data Mining and Knowledge Discovery. 6th European Conference, PKDD 2002, Helsinki, Finland, August 19–23, 2002, Proceedings*: 15-27. Berlin, Heidelberg, Germany: Springer. https://doi.org/10.1007/3-540-45681-3_2.

Brys, G., M. Hubert, and A. Struyf. 2004. "A Robust Measure of Skewness". *Journal of Computational and Graphical Statistics*, Volume 13, Issue 4: 996-1017. https://www.jstor.org/stable/27594089.

Campos, G.O., A. Zimek, J. Sander, R.J.G.B. Campello, B. Micenková, E. Schubert, I. Assent, and M.E. Houle. 2016. "On the evaluation of unsupervised outlier detection: measures, datasets, and an empirical study". *Data Mining and Knowledge Discovery*, Volume 30: 891-927. https://doi.org/10.1007/s10618-015-0444-8.

Cortes, D. 2022. *isotree: Isolation-Based Outlier Detection. R package version 0.5.15*. https://CRAN.R-project.org/package=isotree.

de Waal, T., J. Pannekoek, and S. Scholtus. 2011. *Handbook of Statistical Data Editing and Imputation*. Hoboken, NJ, U.S.: John Wiley & Sons.

D'Orazio, M. 2022. *univOutl: Detection of Univariate Outliers. R package version 0.3*. https://CRAN.R-project.org/package=univOutl.

Ester, M., H.-P. Kriegel, J. Sander, and X. Xu. 1996. "A density-based algorithm for discovering clusters in large spatial databases with noise". In Simoudis, E., J. Han, and U. Fayyad (Eds.). *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96)*: 226–231. Washington, D.C., U.S.: AAAI Press.

Eurostat. 2014. *Memobust Handbook on Methodology of Modern Business Statistics*. Luxembourg: Eurostat. https://ec.europa.eu/eurostat/cros/content/handbook-methodology-modern-business-statistics_en.

Hautamäki, V., I. Kärkkäinen, and P. Fränti. 2004. "Outlier Detection Using k-Nearest Neighbour Graph". In *Proceedings of the 17th International Conference on Pattern Recognition - ICPR 2004*: 430-433. New York, NY, U.S.: IEEE.

Hahsler, M., and M. Piekenbrock. 2022. *dbscan: Density-Based Spatial Clustering of Applications with Noise (DBSCAN) and Related Algorithms. R package version 1.1-10.* https://CRAN.R-project.org/package=dbscan.

Hahsler, M., M. Piekenbrock, and D. Doran. 2019. "dbscan: Fast Density-Based Clustering with R". *Journal of Statistical Software*, Volume 91, Issue 1: 1-30. https://doi.org/10.18637/jss.v091.i01.

Hariri, S., M.C. Kind, and R.J. Brunner. 2021. "Extended Isolation Forest". In *EEE Transactions on Knowledge and Data Engineering*, Volume 33, N. 4: 1479-1489. https://doi.org/10.1109/TKDE.2019.2947676.

Hidiroglou, M.A., and J.-M. Berthelot. 1986. "Statistical editing and Imputation for Periodic Business Surveys". *Survey Methodology*, Volume 12, N. 1: 73-83. https://www150.statcan.gc.ca/n1/en/pub/12-001-x/1986001/article/14442-eng.pdf?st=6rMUAGoq.

Hidiroglou, M.A., and N. Emond. 2018. "Modifying the Hidiroglou-Berthelot (HB) method". *Unpublished Note*. Ottawa, Ontario, Canada: Statistics Canada, Business Survey Methods Division.

Hubert, M., and E. Vandervieren. 2008. "An Adjusted Boxplot for Skewed Distributions". *Computational Statistics & Data Analysis*, Volume 52, Issue 12: 5186-5201. https://doi.org/10.1016/j.csda.2007.11.008.

Italian National Institute of Statistics - Istat, Statistics Netherlands - CBS, and Swiss Federal Statistical Office - SFSO. 2007. *Recommended Practices for Editing and Imputation in Cross-Sectional Business Surveys*. Luxembourg: Eurostat, EDIMBUS Project. http://ec.europa.eu/eurostat/documents/64157/4374310/30-Recommended+Practices-for-editing-and-imputation-in-cross-sectional-business-surveys-2008.pdf.

Knorr, E.M., and R.T. Ng. 1998. "Algorithms for Mining Distance-Based Outliers in Large Datasets". In Proceedings of the 24th *International Conference on Very Large Databases (VLDB '98):* 392-403. San Francisco, CA, U.S.: Morgan Kaufmann Publishers Inc. https://www.vldb.org/conf/1998/p392.pdf.

Liu, F.T., K.M. Ting, and Z.-H. Zhou. 2012. "Isolation-based anomaly detection". *ACM Transactions on Knowledge Discovery from Data* (*TKDD*), Volume 6, Issue 1 (Article 3): 1-39. https://doi.org/10.1145/2133360.2133363.

Liu, F.T., K.M. Ting, and Z.-H. Zhou. 2008. "Isolation forest". In *Proceedings of the Eighth IEEE International Conference on Data Mining* (*ICDM*): 413–422. New York, NY, U.S.: IEEE. https://doi.org/10.1109/ICDM.2008.17.

Madsen, J.H. 2018. *DDoutlier: Distance & Density-Based Outlier Detection. R package version 0.1.0*. https://CRAN.R-project.org/package=DDoutlier.

R Core Team. 2022. R: *A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Ramaswamy, S., R. Rastogi, and K. Shim. 2000. "Efficient Algorithms for Mining Outliers from Large Data Sets". *ACM SIGMOD Record*, Volume 29, Issue 2: 427-438. https://doi.org/10.1145/335191.335437.

Srikanth, K. 2021. *solitude: An Implementation of Isolation Forest. R package version 1.1.3*. https://CRAN.R-project.org/package=solitude.

United Nations Statistical Commission, and Economic Commission For Europe - UNECE. 2000. *Glossary of Terms on Statistical Data Editing*. Geneva, Switzerland: United Nations. https://unece.org/DAM/stats/publications/editing/SDEGlossary.pdf.