

BUILDING AN INTEGRATED DATABASE FOR THE TRADE SECTOR FOR THE PERIOD 2010- 2022

Maria Rita Ippoliti, Luigi Martone, Fabiana Sartor¹

Abstract. This paper aims at describing the process of combining and merging data from the retail trade survey with data stemming from the survey on business confidence in retail trade, for the period from 2010 to 2022. While our previous paper focused on macro data, this paper explores the comparability of micro data. Although the samples of the two surveys involved in the study have a different sectoral coverage, combining them interestingly allows checking the consistency between the evaluations expressed by the economic operators in the qualitative survey and the indicators of the quantitative survey. The merged dataset is a valuable resource to study the enterprises' expectations in the retail trade sector as it includes precious information such as retail trade index, economic activity, channel of distribution and size of the enterprises and enterprises' judgements and expectations on sales for the main economic and business variables.

The paper is organised as follows: a general overview of data sources and a description of the key steps for combining the databases of the two mentioned surveys. The following part describes more in detail all actions implemented to build the integrated dataset and analyses the units involved in both samples through descriptive statistics, graphs and statistical tests. Comparing the two samples highlighted that the number of units involved in both surveys has increased over the years and that there is a good correspondence between qualitative data and quantitative data.

¹ Though the article is the result of a joint work, the single paragraphs are attributed as follows: paragraph 1 and 2 to Fabiana Sartor; paragraph 3 to Maria Rita Ippoliti; paragraph 5 and 6 to Luigi Martone. The published articles are exclusively expressing the authors' opinions; Istat shares no responsibility for the published contents.

1. Introduction

In recent years integration of data from different sources in economic analysis has gained a key role as it allows the study of the trend of the Italian production, exploiting the interactions of available sources.

Qualitative data generally provide information on macroeconomics, catching the overall picture. When compared to traditional quantitative data, qualitative variables have the advantage of a better timeliness (Lui et al., 2011a e 2011b).

This study aims at evaluating the consistency among judgments and expectations of the economic agents of trade sector and data on turnover returned by the same agents in quantitative surveys. That allows exploring more in detail the behavior of the operators in trade sector, evaluating any distortions in the answers provided to the qualitative survey.

Although efforts to integrate data from different sources have already been carried out (see Margani and Orsini, 2020), the present study represents the first work concerning the trade sector, integrating micro data from the Retail Trade Survey and the Business Confidence in Retail Trade concerning years from 2010 to 2022 (see Ippoliti M., Martone L., Sartor F. 2021 for macro data integration).

The integrated database, which incorporates information on national sales turnover, economic activity classification, employment size of enterprises, channel of distribution, enterprises' judgments and expectations on main economic and business variables, appears remarkably useful to study overall enterprises' expectations in trade sector.

2. Trade surveys

2.1. Business Confidence Survey in Retail Trade (FIDCOM)

The European Commission coordinates a harmonised project for Member States to provide data on Business Confidence Survey in Retail Trade (FIDCOM). These surveys allow to compare information on the economic evolution of retail trade at European level (NACE Division G, except for Division 46 - Wholesale trade, except of motor vehicles and motorcycles and for Group 47.9 - Retail trade not in stores, stalls or markets including retail sales via mail order or via Internet).² The survey asks enterprises to express their opinions (judgements and expectations over the following 3 months) about the main economic variables (orders placed with

² Divisions of NACE Section G involved in the survey are Division 45 (Wholesale and retail trade and repair of motor vehicles and motorcycles) and Division 47 (Retail trade, except of motor vehicles and motorcycles).

suppliers, employment, selling prices), giving therefore an updated overview on the evolution of the sector and of the perceived economic uncertainty. Respondents are requested to state their consideration on their total sales in the last three months, on their current volume of stock and on prices charged by their suppliers. Additional questions are asked to know their expectations for the following three months on volume of orders, employment, prices they charge and on total sales. The nominal sample of the Business Confidence Survey in Retail Trade comprises about 1.000 commercial enterprises. Three stratification criteria are used: enterprise employment size class (1-2 employees, 3-5; 6-999; at least 1.000 employees), geographical area (North-West, North-East, Centre, South and the Islands) and main activity (45.1 sales of motor vehicles; 45.2-45.4 maintenance of motor vehicles and sales of accessories; 47.1, 47.2 retail sales of food, drinks and tobacco; 47.3 retail sales of automotive fuel; 47.4-47.7 retail sales of other goods). Enterprises with less than 1.000 employees are randomly sampled, while all units with more than 1.000 employees are included in the sample. The data processing method sets out the estimate of the frequency percentages of each reply option relating to each item of the questionnaire. For this purpose, the processing of the micro data is based on a double weighting system: a) the frequencies of each reply option are firstly weighted using the number of employees declared by the enterprise at the time of the interview (internal weight); b) subsequently fixed weights reflecting the distribution of the added value of the reference sector (external weight) are used. Since March 2015, the aggregation procedure uses an external weighting structure derived from the added value at factor cost referred to 2012. Each variable is measured calculating balances as percentage differences between favourable and unfavourable responses. Weighted balances are seasonally adjusted if needed. The Index of Business Confidence in Retail Trade is calculated as the arithmetic mean of seasonally adjusted balances based upon opinions and expectations on sales and upon judgments on volume of stocks (the above-mentioned values have inverse signs).

2.2. Retail Trade Survey (VEN)

The Retail Trade Index (VEN) produces a short-term indicator measuring the changes in the value and volume of sales. The reference population of the survey is all enterprises having retailing as their main economic activity, except retail trade of motor vehicles and motorcycles and automotive fuel. Therefore, the survey covers the retail trade sector only partially (NACE Rev. 2, G 47 - Retail trade, except of motor vehicles and motorcycles not including automotive fuel)³. Estimates of Retail

³ According to NACE Rev. 2, Retail trade (Division G47) is first classified by type of sale outlet (retail trade in stores: groups 47.1 to 47.7; retail trade not in stores: groups 47.8 and 47.9). For retail sale in

Trade Survey provide a timely indicator of economic performance and strength of consumer spending. Monthly indices on retail trade are released at national level, consistently with the European Union Regulations concerning short-term statistics (see European Regulations n. 1165/98 and n. 1158/2005)⁴. The sample of the survey involves over 8.000 enterprises, which are resident in Italy.

The sample is stratified considering the following variables: main activity according to NACE Rev. 2 and employment size class (1-5, 6-49 and at least 50 employees). The sample includes all large retailers (at least 50 employees) and a representative sample of smaller enterprises. Every year, enterprises belonging to the sample strata of employment size class 1-5 and 6-49 are partially replaced either if they stayed in the sample for more than 3 years, or if they changed their main economic activity or if they closed.

The sampling design of the survey rotates some units out and rotates new units in each year (belonging to employment size classes 1-5 and 6-49 only) to share burden and refresh the sample. Typically, every year 2.500 to 3.000 enterprises are replaced, therefore 60% of the sample does not rotate. This aspect gains relevance when comparing retail trade indicators with qualitative data, as the Retail Trade Survey struggles to keep track of quick evolution of stores' closures and openings. According to their distribution channel, enterprises in the retail trade sample can be classified into large-scale distribution, small-scale distribution, internet sales and non-store sales. Within the weighting structure of Monthly Retail Trade Index (base=2015), large scale-distribution accounts for 46.4% of total turnover, while small-scale distribution reaches 48.0% of total turnover. Retail trade indices are calculated as weighted means of the sub-indices of each stratum. Concerning the calculation method of the indicator, the synthetic index numbers are constructed as weighted averages of indices related to the domains identified by the intersection of the stratification variables (main activity and employment size). The Laspeyres index is used to calculate aggregate indices up to the retail trade total. The weights are based on turnover data from SBS of the year 2015. Value of sales indices measures the retail trade turnover over time at current prices and, therefore, incorporates the effects in changes of quantity sold and prices. In order to determine estimates on the volume of sales, value of sales indices are processed to allow removing price effects on turnover, using the Harmonised index of consumer prices (HICP). Monthly data are first revised in the following month after publication (which occurs 38 days past the reference period). On annual basis, provisional indices are subject to a second revision to calculate the finale estimates.

stores, there exists a further distinction between specialised retail sale (groups 47.2 to 47.7) and non-specialised retail sale (group 47.1).

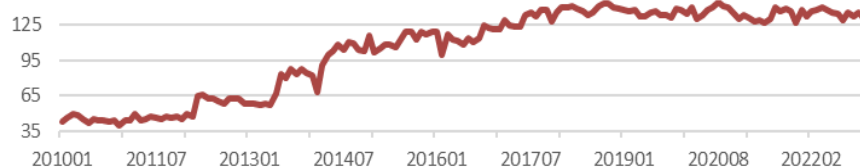
⁴ See <https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CONSLEG:1998R1165:20120621:EN:PDF>

3. The integrated dataset: building and analyses

As a first step, the dataset of Business Confidence in Retail Trade (FID.COM) was extracted, then it was merged with the microdata from Retail Trade Survey (VEN). ASIA code (an identifying code of each enterprises according to the statistical archive of active businesses) and reference period (YEARMONTH) were both used as key variables for merging the two datasets.

Figure 1 shows the movement in the total amount of businesses involved in both surveys (retail trade and confidence in retail trade) by year. From the graph, it is clear that the number of these enterprises grew during the considered period, going from 43 at the beginning of 2010 to around 130 at the end of 2022, with a significant increase starting from the first months of 2014, when the Statistical Portal of Enterprises was launched by Istat for data collection purposes. The integrated VEN.FIDCOM dataset contains all micro data from the Retail Trade Survey and the Business Confidence in Retail Trade for each enterprise and for each month of the year, even if the enterprise is involved in only one of the two surveys. Furthermore, it is worthwhile noting that within the integrated database, in addition to the above—mentioned variables, for each enterprise we calculated the year-on-year growth in turnover. This latter variable is equal to null when the value of turnover is missing in one of the considered periods.

Figure 1 – *Enterprises responding both to the Retail Trade Survey and to the Business Confidence in Retail Trade Survey (years 2010-2022).*

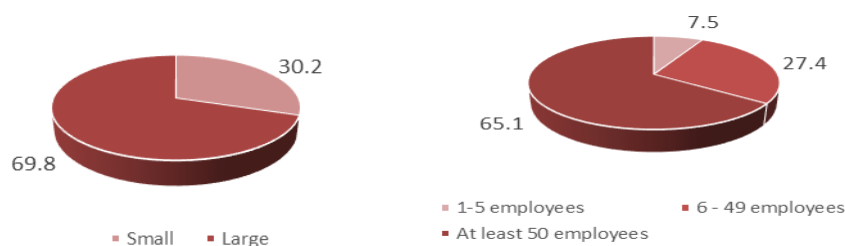


Source: Elaboration on ISTAT data

Figure 2 shows that the number of enterprises responding to both the Retail Trade Survey and the Business Confidence in Retail Trade Survey mostly belong to large-scale distribution (69.8% against 30.2% of enterprises belonging to small distribution).

In order to have more reliable information on the size of the enterprises involved in the study, we chose to consider the number of employees. As a result, the graph on the right of the Figure 2 shows that the amount of large enterprises in the overlapping database is higher (65.1%) when compared to small (7.5%) and medium-sized enterprises (27.4%).

Figure 2 – Percentage of responding enterprises by channel of distribution and percentage of responding enterprises by employment size class in the overlapping database.



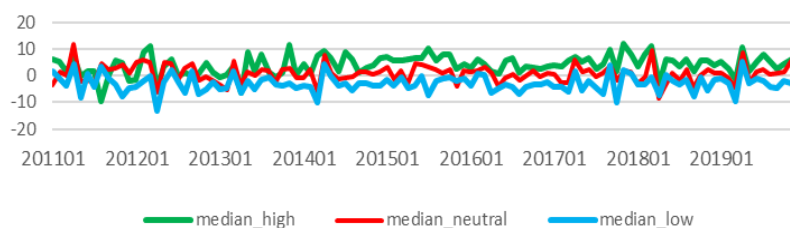
Source: Elaboration on ISTAT data

Interestingly not only descriptive statistics, but also graphical analysis are effective to examine the movement in the year-on-year growth rate of turnover in comparison with the responses to Business Confidence Survey on the evolution of the economic trend. Enterprises involved in both surveys were classified according to their evaluation of the retail trade movement (confidence). For each month three groups of enterprises were created: a group answering "High", a group answering "Neutral" and a group answering "Low"⁵.

Then, for each enterprise and for each month, we calculated the year-on-year growth rate of turnover using micro data from Retail Trade Survey. Subsequently, we made a graphical analysis of the median of the year-on-year growth rates for each response group ("High", "Neutral" and "Low").

Median was preferred over mean because the value of the mean can be distorted by the outliers. In fact, since the year-on-year growth rates are calculated for each enterprise, either really high or really low values are the likely outcomes in certain months.

⁵ The question asked is "How have your SALES developed over the past 3 months?" Responses are "Sales have: 1=increased; 2=remained unchanged; 3=decreased."

Figure 3 – Graph of the median: years 2011-2019.

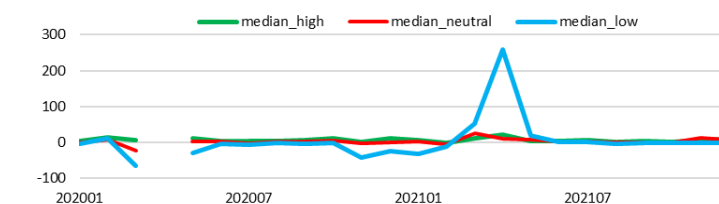
Source: Elaboration on ISTAT data

The median of the "High" response in the judgment on the sales trend generally shows higher values when compared to the median of the "Low" response. Furthermore, the graph reveals that the median of the year-on-year growth rate of the "Neutral" response is almost always placed between high and low. In particular, from 2016 the graph shows a bigger distance among the series because of a higher amount of enterprises involved in both surveys. The graphical analysis, therefore, shows a good alignment between qualitative and quantitative data.

3.1. Focus on 2020-2022

To evaluate the effect of pandemic on data, series were split: 1st period from 2010 to 2019 and 2nd period from 2020 to 2022 (see figure 4). Business Confidence data were not collected in April 2020 because of the pandemic emergency, therefore the graphs detects a discontinuity in the series. Consequently, as the year-on-year growth rates were analysed, the whole year 2021 appears to be affected by the pandemic too.

In particular, in April 2021 the enterprises giving the "Low" response still have a high year-on-year growth rate as they were closed in April 2020.

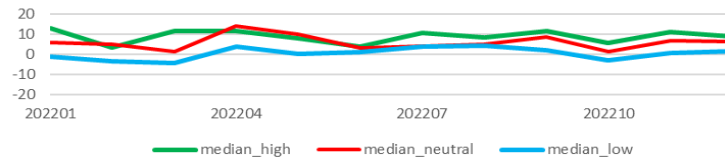
Figure 4 – Graph of the median : years 2020-2021.

Source: Elaboration on ISTAT data

From 2022 the pandemic effect on data seems to fade (see figure 5). In the second semester of the year, the value of the median of year-on-year growth rates related to

the "Neutral" response is almost always placed between the "High" and "Low", just like in the pre-pandemic years.

Figure 5 – Graph of the median: years 2022.

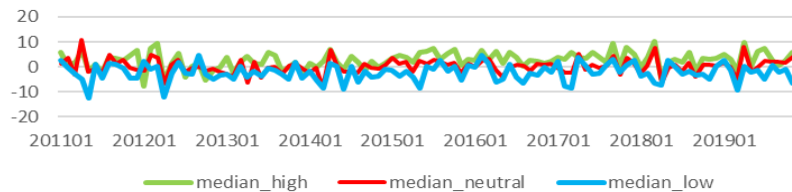


Source: Elaboration on ISTAT data

4. Forecast analysis and Median test

Experts often wonder whether qualitative surveys can be used to forecast quantitative data. For this purpose, we calculated the percentages of enterprises declaring an increase/invariance/decrease in sales in connection to the business trend expectations of the Business Confidence, in order to verify if there was an alignment between the enterprises' forecasts and actual values declared in the following months⁶.

Figure 6 – Graph of median forecast: years 2011-2019.



Source: Elaboration on ISTAT data

Graphs at time t (see figure 6) seem to confirm a good alignment between the two surveys. Therefore this alignment between the two surveys can also be used for forecasting purposes since the median associated with the "High" response of the judgment on sales forecasting, in most cases has higher values than those recorded for the "Low" response.

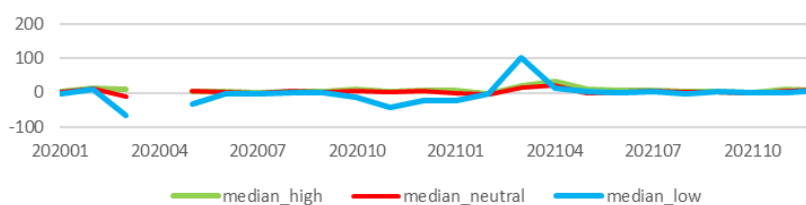
⁶ The question asked is: "How do you expect your SALES to change over the next 3 months?" Responses are "Sales will: 1=increase; 2= remain unchanged; 3=decrease."

4.1. Focus on 2020-2022

To allow forecast analyses, time series were split: 1st period from 2010 to 2019 and 2nd period from 2020 to 2022. Figure 7 shows a discontinuity in the time series due to the lack of Business Confidence data in April 2020.

Business expectations in 2021 were affected by the pandemic emergency as in April 2021 the enterprises that chose the "Low" response have a high year-on-year growth rate in turnover as they were closed in April 2020.

Figure 7 – Graph of median forecast: years 2020-2021.

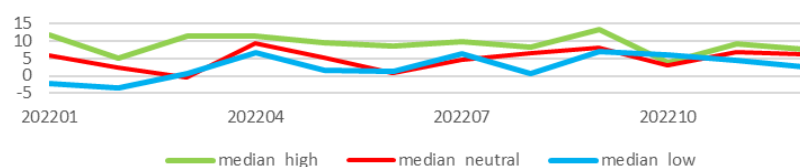


Source: Elaboration on ISTAT data

During pandemic, enterprises struggled to define their business expectations.

Starting from 2022 (see figure 8), especially from June, the value of the median of the year-on-year growth rate for the "Neutral" response mostly falls between "High" and "Low", as it happened in the pre-pandemic years.

Figure 8 – Forecast graph of the median : years 2020-2021.



Source: Elaboration on ISTAT data

4.2. Median test

In addition to descriptive statistics and graphical analysis, statistical tests were carry out on the medians of the year-on-year growth rates at time t for the "High", "Neutral" and "Low" responses for the qualitative variable "Sales trend", starting from the dataset of enterprises involved in both surveys.

Enterprises involved in both surveys were sorted according to the way they answered to the item regarding sales performance (confidence). Then, three groups of enterprises were created for each year: the group that answered "High", the group that answered "Neutral" and the one that answered "Low".

Hence, for each enterprise and for each year, we calculated the year-on-year growth rate of turnover, using data stemming from Retail Trade Survey. We used the non-parametric Wilcoxon test, which verifies the hypothesis that two samples are drawn from populations with coincident medians.

Table 1 – Median and median test (year 2011-2022).

YEAR	MEDIAN			TEST ON MEDIAN(P-VALUE)		
	High	Neutral	Low	High – Low	High – Neutral	Neutral – Low
2011	3.4	1.0	-2.4	0.000	0.204	0.000
2012	0.9	1,7	-3.4	0.010	0.477	0.000
2013	0.5	-0.2	-3.3	0.000	0.400	0.000
2014	5.0	0.2	-3.2	0.000	0.001	0.000
2015	6.3	1.7	-1.8	0.000	0.000	0.000
2016	3.8	0.8	-2.6	0.000	0.000	0.000
2017	5.7	0.9	-2.5	0.000	0.000	0.000
2018	4.5	-0.6	-2.7	0.000	0.000	0.041
2019	4.3	1.3	-2.2	0.000	0.000	0.000
2020	6.8	3.1	-5.6	0.000	0.000	0.000
2021	5.3	4.7	-0.3	0.001	0.962	0.004
2022	8.8	6.1	0.6	0.000	0.000	0.000
2011-2022	5.4	1.5	-2.5	0.000	0.000	0.000

Source: Elaboration on ISTAT data

Test results confirm a good alignment between the two surveys. Apart from year 2012, the alignment of the median of the year-on-year growth rate for the different groups was verified, i.e. the median of the year-on-year growth rates of the "High" group is always higher than that of the "Neutral" group, which is higher than that of the "Low" group. In the very first years and especially when comparing the "High-Neutral" groups, tests accept the null hypothesis of equality of the medians.

This is mainly due to the smaller amount of enterprises involved in both surveys in the first years considered, when compared to the last period. Year 2021, on the other hand, remains a critical year because, as shown also by the graphic analysis, it displays a particular trend due to the pandemic.

5. Conclusions

Analyzing the two survey samples highlights that the number of units involved in both surveys has increased over the years and it shows a good alignment between qualitative and quantitative data.

The graphical analysis of the median of the year-on-year growth rates of turnover in connection to the responses to the Business Confidence Survey confirms a good alignment between qualitative and quantitative data.

The forecast analysis shows a good alignment of the two surveys, since business expectations on turnover recorded by the Business Confidence Survey appear to have a similar pattern of the turnover declared by the same enterprises in the quantitative survey. Findings of the graphical analysis appear to be well substantiated by the results of statistical tests on the median of the year-on-year growth rates at time t for the “High”, “Neutral” and “Low” responses of enterprises involved in both surveys. Apart from a discrepancy recorded in the first years because of a small amount of enterprises involved in both surveys, the results of the tests confirm a good alignment between the two surveys.

References

- LUI S., MITCHELL J., WEALE M. 2011a. Qualitative Business Surveys: Signal or Noise? *Journal of the Royal Statistical Society*, Vol. 174, No. 2, pp. 327-348.
- LUI S., MITCHELL J., WEALE M. 2011b. The Utility of Expectational Data: Firm-Level Evidence Using Matched Qualitative-Quantitative UK Surveys, *International Journal of Forecasting*, Vol. 27, No. 4, pp. 1128-1146.
- MARGANI, P., ORSINI D. 2020. L'indagine sulla Fiducia delle Imprese Manifatturiere e l'indagine sul Fatturato e Ordinativi dell'industria a Confronto: Un Dataset Integrato sulle Imprese Italiane del Comparto Manifatturiero negli Anni 2005-2019. *Working papers 5/2020*, Roma: Istituto Nazionale di Statistica.
- IPPOLITI M., MARTONE L., SARTOR F. 2021. Trade Surveys: Qualitative and Quantitative Indicators, *Rivista Italiana di Economia, Demografia e Statistica*, Vol.75, No 4.

Maria Rita IPPOLITI, Istat, marippoliti@istat.it

Luigi MARTONE, Istat, martone@istat.it

Fabiana SARTOR, Istat, sartor@istat.it

