Cross Cutting Features of the Italian Retail Trade Sector

Paper prepared for session 5 - OTHER CROSS CUTTING ISSUES

Paola Anitori & Roberto Gismondi[1] *- ISTAT - ITALY*

ABSTRACT

This paper contains some thoughts about the informative content of the ATECO 91 classification - actually used in ISTAT - referring in particular to the retail trade sector (Division 52). The main aim is to verify if the recourse to other classifications more aggregated, or in same way complementary to ATECO, could contribute to maintain steady or to increase the discriminant level of the classification itself. With these premises, the paper goes on with paragraph 2, in which a cross-cutting analysis based on breakdown of turnover by product is presented. Paragraph 3 attempts to provide a rough evaluation of the not observable heterogeneity implicit when using NACE or the Italian ATECO classification. In paragraph 4 a discriminant model is tested in order to identify which variables, measured at the enterprise level, are able to explain different levels of retail turnover. A final short paragraph tries to summarise the main results, raising some questions to be discussed.

1. Premise: some limits of NACE Rev.1 classification for Division 52

The actual main EU classification of activities (NACE Rev.1) stratifies retail trade enterprises according to their "prevalent activity", on the basis of the typology of products, i.e. the "physical" characteristics of the goods sold and, but only to a lower extent, the type of sale-service offered to clients, mostly limited to the specification of the type of outlet in which sales have been done.

Even though NACE is a simple, homogeneous and consistent tool to harmonise data capturing, classification and dissemination, the actual state of the art on that in Italy suggests some reflections about the need to evaluate more in deep the real correspondence between this classification and the real retail trade environment. Some weak points of NACE seem to be the following:

- *it doesn't highlight the overlapping between type of product and the type of sale service both affecting the overall sale activity*. The most remarkable example, other than the film processing service provided by optical and photographic equipment outlets, is represented by all the services related to the electronic and computer equipment sales supplied by specialised shops (Gismondi, 1998.1); from this point of view the international classification is still incomplete since retailers are not able to shift the sale activity from other sale services they provide to clients (including after sale services).
- *Misclassification problems often occur*. It is not seldom to find stores (especially not-food stores) that according to their size and sale area are classified among small retailers, even if the

---

products sold are shown to the clients according to some characteristics of the large-scale distribution such as self-service, counters at the exit, and so on. Moreover, the belonging to any associative form (buying groups, voluntary chains, co-operatives, franchising) could hardly stress this problem because associated enterprises seem to perform quite differently from the others (Istat, 1998.1). We remind that the associative forms are totally excluded from the existing classifications, being considered as groups of enterprises.

- *The concept of "specialisation" is sometimes not realistic.* It is a matter of fact that the split by prevalent activity derived from NACE Rev.1 can not necessarily reflect the real composition of goods and commodities sold by the enterprise; this is particularly true for mixed and not-specialised enterprises and it complicates performance analyses, especially as far as not-food enterprises are concerned.

Given that the EU Regulation on Short-term Statistics, entered into force at the beginning of July, obliges every EU Country to calculate monthly retail trade indexes on the basis of the prevalent activity of enterprises (see Eurostat, 1998.1 and Gismondi, 1998.2), it's straightforward to evaluate the rightness and the realism of this principle, strongly linked to the concept of specialisation and hardly affected by the possible misclassifications above mentioned.

With these premises the paper goes on with paragraph 2, in which a cross-cutting analysis based on breakdown of turnover by product is presented. Paragraph 3 attempts to provide a rough evaluation of the not observable heterogeneity implicit when using NACE or the Italian ATECO classification. In paragraph 4 a discriminant model is tested in order to identify which variables, measured at the enterprise level, are able to explain different levels of retail turnover. A final short paragraph tries to summarise the main results, raising some questions to be discussed.

2. A preliminary cross-cutting analysis based on breakdown of turnover by product

A particular and simple classification, thought for the monthly retail sales survey, is currently used by ISTAT in addition to NACE Rev.1 (and to ATECO 91, the Italian classification equivalent to NACE up to four digits), that remains the basic classification from which this special one is derived. This *ad hoc* classification is based on 15 groups of products (plus three additional groups for not specialised stores and sales by mail, as it will seen more in deep in paragraph 3) and is the result of contacts with category associations, users, traders, research centres, national accounts experts. It can be considered more suitable for the users' needs and able to guarantee an easier compilation for enterprises, the most part of which uses own classifications for products sold[2]. In this context we will not examine the enterprises performing retail sales activities *only* in a form different from the primary, but the possible multiple retail sales activities characterising enterprises classified as operating in the NACE specialised groups 52.2, 52.3, 52.5, or in the class 52.61.

Why such analysis? The basic idea is that when surveying the retail trade sector using an approach *by enterprise*, as it's requested in the above mentioned Short-term Business Statistics Regulation, we must remember that, generally speaking, an approach by enterprise *does not eliminate the arbitrary element in the modes of utilising the data captured.* For instance, the questionnaire on monthly sales is normally

---

[2] We underline the correspondence between the 15 groups and the 5-digit ATECO classifications: **1** = 52.2 (without 52.26) - 52.11.3 – 52.11.4; **2** = 52.31-52.32; **3** = 52.41 (without 52.41.2) - 52.42; **4** = 52.43; **5** = 52.44 (without 52.44.2 and 52.44.5) - 52.41.2; **6** = 52.45.1; **7** = 52.45.2 - 52.48.1; **8** = 52.48.2; **9** = 52.44.2 - 52.44.5; **10** = 52.46; **11** = 52.33; **12** = 52.47; **13** = 52.45.3-52.45.4; **14** = 52.48.4-52.48.5; **15** = 52.45.5-52.48.3-52.48.6-52.48.9.

addressed to the enterprise, that is the relevation unit, but it may be designed to capture the value of sales for each different kind of product sold as well, as in the Italian case. As a consequence, we have to face the choice of how calculating the monthly sales indexes: on the basis of the enterprise's predominant activity only (each enterprise is used only once for the calculations), or utilising the same enterprise as many times as the kinds of products in relation to which not-null sales have been registered in the form? The choice depends strictly on the degree of correspondence between the enterprise's main activity resulting from the available extraction list, and on the degree of precision with which the prevalent activity fully expresses the actual typological array of products sold. *As a rule, the choice leads to different results*, and evidences on that were reported in Ceccarelli, Gismondi and Mirto (1997).

On the basis of a classification that is more functional than typological, the same enterprise, when selling highly different types of products, should be counted as many times as the different consumption functions dealt with. In this regard, Eurostat has been trying for some time to promote among the member States application of the CPA classification (based on the "type of sale-service" rather than on the type of product marketed), together with the NACE Rev.1 classification.

Main results are showed in table 2.1. They must be considered as provisional, because a more exhaustive project referred to this matter and concerning 1997 data is currently going on[3]. About 8.000 specialised enterprises responding for at least 9 of the 12 months of 1996 declared in the questionnaire the split of their incomes in correspondence of groups of products, eventually additional to the group in which the enterprise was originally classified on the basis of its prevalent activity resulting from the extraction list. In the table we inserted the percent share of enterprises selling goods corresponding to groups different from the prevalent one, using NACE at 3 (3 groups) and 4 digits (17 classes were covered by the sample[4]), the special 15 groups ISTAT classification and the 5 specialised groups of products requested in the frame of the Short-term Regulation[5]. A 100% value means that all the (sample) enterprises with a certain prevalent activity sell goods corresponding to only one group of products. The position of each group according to the decreasing ranking of the previous percentage is provided as well (in brackets), so that the value "1" identifies the group for which the concept of specialisation seems to be less relevant in comparison with the others.

First of all, we must point out that different overall percentages can be derived varying the classification considered, because every average percentage is obtained by a weighted arithmetic mean of the correspondent percentages by groups. So we have that:

- using 3 digits NACE only the 0,4% of enterprises sell goods corresponding to groups different from the prevalent;
- using 4 digits NACE the percentage is equal to 8,3 (clearly we are referring to classes);
- using the ISTAT classification based on 15 groups the percentage is equal to 9,8;
- using the 5 specialised groups requested by EUROSTAT the percentage raises to 10,5.

In other words the 4 digits NACE seems to be the most reliable classification (the 3 digits NACE is too raw to be useful for describing retail trade different behaviours), but the ISTAT classification (including 15 groups instead of 17) is rather good as well. On the other hand the 5 specialised groups requested by EUROSTAT seem to be a bit more raw.

Moreover, the classification based on the 4 digits NACE shows even more clearly how the

---

[3] For more details see Eurostat (1998.1).
[4] These 17 NACE classes correspond to 53 ATECO 91 categories.
[5] The EUROSTAT eight groups are identified by the following NACE groups and classes: 52.11; 52.12; 52.2; 52.3; 52.41-52.43; 52.44-52.46; 52.47-52.48; 52.61. For further details see Gismondi (1998.2).

prevalence guideline is less effective for all the classes from 52.44 to 52.48, as well as for 52.32, relative to "Medical and orthopaedic articles" and, to a lesser extent, 52.24, relative to "Bread, pastries and sweets". For all the other classes the guideline can be considered quite satisfactory.

Considering the 15 ISTAT groups and the 5 EUROSTAT ones, it's evident that the four national groups with a share of enterprises with sales only in correspondence to the pertinent group below 80% are (column 1): "Toys, games, sports, camping articles" (64,8%), "Household appliances" (65,1%), "Radio, television, recorders, information technology products" (74,3%), and "Ironmonger's articles" (79,0%). With reference to the Eurostat classification, the foregoing circumstance occurs only for "Furniture and articles for the house" (72,2%); on the other hand, groups of products better specified by their own prevalent activity are: "Leather products" (95,4%) and "Photographic goods" (94,8%) for the national grouping, and "Food and beverages" (also 94,8%) for both ISTAT's and EUROSTAT's classifications.

Table 2.1 - Precision of the "specialisation" concept for some Italian retail trade firms

Code	Products	% sales in other groups	Rank	Mean	Code	Products	% sales in other groups	Rank	Mean
	3 digits NACE					**15 ISTAT groups**			
522	Food and beverages	0,7	(1)	0,4	1	Food and beverages	5,2	(13)	9,8
523	Pharmaceuticals - cosmetics	0,0	(3)		2	Pharmaceutical products	10,2	(10)	
524	Other products	0,5	(2)		3	Clothes and shoes	6,8	(12)	
					4	Leather products	4,6	(15)	
	4 digits NACE				5	Furniture, textiles for the house	12,7	(8)	
5221	Fruit and vegetables	4,6	(13)	8,3	6	Electrical household appliances	34,9	(2)	
5222	Meat	3,4	(15)		7	Radio, tv, personal computers	25,7	(3)	
5223	Fishes	0,0	(17)		8	Photographic goods	5,2	(13)	
5224	Bread and cakes	12,8	(7)		9	Small products for the house	19,8	(5)	
5225	Beverages	2,8	(16)		10	Ironmonger's shops	21,0	(4)	
5227	Other food products	6,2	(11)		11	Cosmetics and personal goods	10,3	(9)	
					12	Stationery, books, newspapers	17,1	(7)	
5231	Pharmaceutical products	9,7	(9)		13	CD, tapes, musical instruments	10,0	(11)	
5232	Orthopaedic appliances	15,4	(6)		14	Toys, games, sport, camping	35,2	(1)	
5233	Cosmetics	10,3	(8)		15	Other products	17,6	(6)	
5241	Textiles	5,8	(12)			**5 EUROSTAT specialised groups**			
5242	Clothes and shoes	6,4	(10)		E3	Food and beverages	5,2	(5)	10,5
5243	Leather articles	4,6	(13)		E4	Pharmaceuticals, cosmetics	10,3	(3)	
5244	Furniture and lighting articles	19,5	(3)		E5	Clothes and shoes	5,7	(4)	
5245	Electrical household appliances	29,0	(1)		E6	Furniture, articles for the house	27,8	(1)	
5246	Ironmonger's shops	28,2	(2)		E7	Other products	18,8	(2)	
5247	Stationery, books, newspapers	17,1	(5)						
5248	Other products	18,1	(4)			**TOTAL**	**17,0**		

Source: elaboration on ISTAT data (1996).

The questionnaire used by ISTAT for the monthly retail trade survey is able to break down sales by type of outlet as well. The typologies are the following: small and medium size "traditional" shops (more or less all specialised), hypermarkets, supermarkets, minimarkets, department stores, hard discounts, other big specialised shops; mail sales are investigated too (NACE 52.61). The relevance of this information consists in the possibility to evaluate the degree of precision of the classification from another point of view: without taking into account the type of product sold, we want to verify if products have been sold in only one type of outlet; in other words, the matter is to detect enterprises operating with outlets classified in branches different from the one in which the enterprise itself is classified as mainly operating (for instance, an enterprise classified in the category

52.11.2 "supermarket" could operate with some minimarkets as well, included in the category 52.11.3).

Table 2.2 shows that, on the average, the 96,0% of retail enterprises operate with one type of outlet only (the one corresponding to the main activity), whilst the 3,7% operate with two types of outlets (like the previous example) and only the 0,3% with more than two types of outlets.

As it is reasonable, on of the lowest shares of enterprises operating with one type of outlet only characterises not specialised shops (NACE 52.1 with the 91,7%), even though the overall lowest share concerns a specialised NACE class (52.25, sales of beverages, with the 90,0%).

Generally speaking, the breakdown by type of outlet, that could be intended as a particular "type of sale service" offered to the client, doesn't seem so relevant as the breakdown by product seen in table 2.1, meaning that *coeteris paribus* the heterogeneity of retail enterprises seems to depend more on the type of product than on the type of outlet in which these products are sold.

In any case, it's quite clear that the relation between type of product sold and trading form (type of outlet), implicitly set by the classification, is becoming less relevant than in the past, since deep changes in clients' needs imposed a strong and quick transformation of the retail trade sector.

Table 2.2 - Precision of the "type of outlet" concept for some Italian retail trade firms

Code	Products	% of cases with one type	% of cases with two type	% of cases with > 2 type	Code	Products	% of cases with one type	% of cases with two type	% of cases with > 2 type
	3 digits NACE (including 52.6)					**15 ISTAT specialised groups**			
521	Not specialised sales	91,7	6,9	1,4	1	Food and beverages	97,2	2,7	0,1
522	Food and beverages	97,0	2,9	0,1	2	Pharmaceutical products	96,6	3,4	0,0
523	Pharmaceuticals, cosmetics	96,8	3,2	0,0	3	Clothes and shoes	96,1	3,9	0,0
524	Other products	96,7	3,3	0,0	4	Leather products	97,1	2,9	0,0
526	Sales outside shops	98,4	1,6	0,0	5	Furniture, textiles for the house	95,4	4,6	0,0
					6	Electrical household appliances	96,8	3,2	0,0
	4 digits NACE				7	Radio, tv, personal computers	94,9	4,8	0,3
5211	Not specialised food prevailing.	91,8	6,9	1,3	8	Photographic goods	96,4	3,6	0,0
5212	Not specialised not food prevail.	90,6	5,7	3,7	9	Small products for the house	97,4	2,6	0,0
					10	Ironmonger's shops	97,4	2,6	0,0
5221	Fruit and vegetables	96,3	3,7	0,0	11	Cosmetics and personal goods	97,1	2,9	0,0
5222	Meat	97,5	2,5	0,0	12	Stationery, books, newspapers	96,5	3,5	0,0
5223	Fishes	100,0	0,0	0,0	13	CD, tapes, musical instruments	98,8	1,2	0,0
5224	Bread and cakes	98,0	2,0	0,0	14	Toys, games, sport, camping	95,3	4,7	0,0
5225	Beverages	90,9	6,8	2,3	15	Other products	96,3	3,7	0,0
5227	Other food products	96,9	3,1	0,0					
5231	Pharmaceutical products	96,8	3,2	0,0					
5232	Orthopaedic appliances	92,9	7,1	0,0		**5 EUROSTAT specialised groups**			
5233	Cosmetics	97,1	2,9	0,0	Es1	Food and beverages	97,0	2,9	0,1
					Es2	Pharmaceuticals, cosmetics	96,8	3,2	0,0
5241	Textiles	98,0	2,0	0,0	Es3	Clothes and shoes	96,6	3,4	0,0
5242	Clothes and shoes	95,9	4,1	0,0	Es4	Furniture, articles for the house	97,0	3,0	0,0
5243	Leather articles	97,1	2,9	0,0	Es5	Other products	96,1	3,9	0,0
5244	Furniture and lighting articles	96,9	3,1	0,0					
5245	Electrical household appliances	96,7	3,2	0,1					
5246	Ironmonger's shops	97,4	2,6	0,0					
5247	Stationery, books, newspapers	96,5	3,5	0,0		**TOTAL (52.1+52.2+52.3+ +52.4+52.6)**	**96,0**	**3,7**	**0,3**
5248	Other products	96,0	4,0	0,0					

Source: elaboration on ISTAT data (1996).

3. Degree of precision of different breakdowns in Division 52: the empirical evidence

In Italy, the classification ATECO 91 (national version of NACE Rev.1 with which it coincides up to 4 digits) tries to add further details subdividing the 4 digit classes into 5 digits categories but, unfortunately, it replies the same limits of the former so that it needs to be revised as well. Some attempts have been made to find new ways of aggregation of categories or classes of products in order to improve the quality of analyses; some of these aggregations have been suggested by EUROSTAT, others by the users, but the bulk of the discussion relies on the need of reconciling the analysis by industry (allowed for by the classification) with the analysis by product and/or by type of outlet aimed at by the users.

According to the existing sample survey on retail trade (based in 1997 on a panel of about 7.800 enterprises monthly interviewed to calculate turnover indexes), we tried to evaluate the efficiency of the ATECO 91 classification at 5-digit level compared to other aggregations of products in order to draw some conclusions about the need of a possible re-classification of products or extension of the classification.

An analysis of variance was iterated (see table 2.1) and the dependent variable, given by the individual turnover of each enterprise, was analysed according to the levels of the control variable, i.e. according to the categories of the ATECO 5-digit level categories, 4-digit level groups and 3-digit level classes, to the 15 groups of products used in the survey and to the 8 groups required by Eurostat.

Table 3.1 - Analysis of variance according to different groups of products

Variance %	ATECO 5-digit (60 groups)	ATECO 4-digit (21 groups)	ATECO 3-digit (5 groups)	15 ISTAT groups	8 EUROSTAT groups
Explained by the model (R^2)	10,1	4,3	0,3	3,5	4,3
Residual	89,9	85,7	99,7	86,5	85,7
Total	100	100,0	100,0	100,0	100,0
Fisher's F	14,8	17,5	6,2	18,9	50,0
Significance[6]	*0,0001*	*0,0001*	*0,0001*	*0,0001*	*0,0001*

Source: elaboration on ISTAT data (1997).

It is quite clear that the "bottom-up" approach, that allows to pass from the 5-digit level to larger groups such as the 4-digit headings or the 3-digit headings, produces a decreasing trend of R^2, meaning a worsening in the quality of the survey results. In particular, whilst the 4-digit classes and the *ad hoc* groups requested by ISTAT and EUROSTAT seem to produce the same results, the 3-digit groups show to be the less efficient in terms of explained variance, as to confirm that a certain level of detail of products has to be maintained.

However, the value of the residual (i.e. variance not explained by the model) suggests to go more into depth since the frame shown by table 3.1 does not seem to be clear enough to highlight possible inconsistencies in the classification. As a consequence, the coefficients of variation (CV) of individual turnover were calculated according to all the categories of the control variable, in order to detect both those cases where the "bottom-up" approach would allow to restrict the analysis to a "gross" stratification without loosing too much information, and those cases where it would be better to think to a different structure of the classification.

Comparing table 3.2, which shows the coefficients of variation (CV) for each single 5-digit ATECO, with table 3.3, showing CV values related to the other classifications considered in table

---

[6] Probability that F>F*.

3.1, together with the minimum and the maximum of each CV and the corresponding number of enterprises in the universe, it can be said that the "bottom-up" approach is not always statistically efficient since there are cases where, in spite of the low value of the CV shown by the 4-digit heading, the single categories of the heading are very different from one another in terms of variability of turnover.

Table 3.2 - Coefficients of variation of turnover at 5-digit level of ATECO 91 Classification - Division 52 (1997 sample survey)

ATECO	Products	Coefficient of variation	Number of enterprises	% Number of enterprises
52112	Supermarkets	-	4.624	0,8
52113	Minimarkets	5,76	11.663	2,0
52114	Large scale food stores	4,69	64.388	11,2
52121	Department stores	3,03	516	0,1
52122	Large scale not specialised stores	2,41	2.429	0,4
52210	Fruit and vegetables	2,28	23.733	4,1
52221	Cattle, pork, ovine, horse meat	1,85	37.725	6,6
52222	Poultry, game	0,76	3.298	0,6
52230	Fish	1,73	5.999	1,0
52241	Bread	1,23	4.653	0,8
52242	Cakes	2,74	3.915	0,7
52250	Beverages	2,4	4.629	0,8
52260	Tobacco and other related goods	0,92	21.728	3,8
52271	Milk	0,99	3.859	0,7
52272	Groceries	2,3	8.032	1,4
52273	Coffee	2,29	1.593	0,3
52274	Other food and beverages	1,97	2.692	0,5
52310	Pharmaceutical products	1,66	15.107	2,6
52320	Orthopaedic appliances	1,08	2.194	0,4
52331	Herbalists	0,86	2.989	0,5
52332	Perfumeries, cosmetics and personal goods	3,2	12.067	2,1
52411	Textiles for dressing	3,04	3.922	0,7
52412	Textiles for the house	0,94	3.178	0,6
52413	Linen	0,95	3.048	0,5
52414	Hosieries and knitwears	0,74	1.034	0,2
52421	Adults clothes	3,61	51.520	9,0
52422	Children and babies clothes	4,96	6.050	1,1
52423	Lingeries, shirts	1,04	17.142	3,0
52424	Haberdasheries	2,25	16.648	2,9
52425	Fur-coats	0,91	1.050	0,2
52426	Hats, umbrellas, gloves, ties	0,93	802	0,1
52431	Shoes	8,29	17.595	3,1
52432	Travel articles	4,26	3.573	0,6
52441	Furnitures	3,42	17.868	3,1
52442	Glasswares and crockeries	5,75	12.656	2,2
52443	Lighting articles and electric materials	1,16	4.374	0,8
52444	Wood, wicker, plastic articles	0,88	792	0,1
52445	Various articles for domestic use	2,95	2.718	0,5
52451	Electrical household appliances	3,98	9.535	1,7
52452	Radio, tv, personal computers	3,49	3.949	0,7
52453	Records and tapes	4,23	2.770	0,5
52454	Musical instruments	1,86	1.173	0,2
52455	Sewing machines	2,39	723	0,1
52461	Ironmonger's shops	1,83	12.422	2,2
52462	Paints	4,94	4.115	0,7
52463	Hygienic and health articles	1,89	2.677	0,5
52464	Building appliances	1,6	6.557	1,1
52465	Hydraulic appliances	2,64	1.111	0,2
52466	Agricolture machines	7,42	6.288	1,1
52471	New books	2,77	3.511	0,6
52472	Nespapers, magazines	0,55	12.036	2,1
52473	Stationery and office stuffs	2,03	16.121	2,8
52481	Office appliances	1,42	3.353	0,6
52482	Optic, photography, precision instruments	6,92	9.347	1,6
52483	Watches, silver articles and jewelry	4,66	17.248	3,0
52484	Toys and games	4,25	4.300	0,8

52485	Sport, bykes, weapons		2,9	18.993	3,3
52486	Art and cult		5,53	4.093	0,7
52489	Other products		0,99	30.146	5,3
52610	Sales by mail		3,04	681	0,1
		Total	2,74	572.952	100,0

Source: elaboration on ISTAT data (1997). 140.635 enterprises belonging to the universe are not
included because classified in categories not covered by the sample.

Generally speaking, the "bottom-up" approach doesn't explain why within some classes some categories show a so high heterogeneity. For an easy comprehension of the tables, the loss (or gain) of efficiency obtainable by different groupings of products can be evaluated directly comparing the CV values of the groups themselves[7].

Additional comments arise from table 3.2. If we give a look to the classes with a high value of the CV and, within each of these classes, we look for the categories with the highest heterogeneity, it can be seen that the number of enterprises in the corresponding domains of the universe of reference for those categories is very large, sometimes larger than the number of enterprises of some entire 4-digit headings. This result is probably not by chance, since it may state the possibility that, beyond the misclassification errors typical of every business register, these categories are too vaguely defined and that there could be the need of more efficient classification rules to take account for the aspects discussed above.

Class 52.42 "Retail sale of clothing" is a quite remarkable example that summarises all it has been said: table 3.3 shows a quite high CV value for the entire class, probably due to the importance of categories 52.42.1 ("Adult clothes", including the 55,3% of enterprises belonging to the class 52.42) and 52.42.2 ("Children and baby clothes") in terms of heterogeneity (see table 3.2), so that it's not advisable to restrict the analysis to the 4-digit heading.

It's more than a doubt that the source of the high internal heterogeneity of these categories is not a chance: table 3.4 shows that the 3,7% of enterprises in the category 52.42.1 and the 19,2% in 52.42.2 (the 5,9% of the two categories together) declared to have at least one outlet running the business in one of the most common associative forms[8].

By the way, table 3.4 shows that the percentage of respondents who declared to have outlets running the business in a different trade form from the ones set by the classification is not negligible: they are the 7,2% over the total and, at 4-digit level, the average percentage per class is around 6,5%. In terms of turnover it concerns the 5,8% of the sample total as to prove that the matter is not of lower importance.

Similar comments can be made for other classes such as 52,44 "Retail sale of furniture, lighting equipment and household articles" and some of the categories regarding food products, in particular 52.11.4 "Other retail trade of food in non-specialised stores", which refers to large-scale distribution with food predominating and that seems more a sort of container where all the enterprises not elsewhere classified are collected. Moreover, it has to be remembered that, at the time being, hard discounts have not yet been classified.

Thus, there is the need of better detecting these categories in order to check if their internal etherogeneity is due, a part form the natural bias due to sampling, to differences among enterprises

---

[7] As a matter of coherence, the CV values performed by EUROSTAT's aggregation should be compared with those of the 4-digit heading (which, by the way, results to be more efficient even though table 3.1 shows the same R^2 value for both the aggregation), whilst the 15 groups concerning ISTAT's monthly survey can be compared directly with all the other aggregations.
[8] Voluntary chains, buying groups, franchising, localisation in a trading centre.

(such as structural differences, number of outlets and so on) or to something else such as, for instance, having sale outlets running the business according to one of the associative forms. If the latter was the case, it would be a first evidence that the classification should be based on different or additional rules, in order to reduce problems when interpreting the economic significance of statistical results.

Table 3.3 - Variability indexes concerning ATECO 91 classification
Division 52 (1997 sample survey)

ATECO	CV	LOWER ORDER CLASSES		
		CV MEAN (1)	CV MIN (2)	CV MAX (2)
4-DIGIT ATECO				
5211	5,67	5,23	4,69	5,76
5212	4,72	2,72	2,41	3,03
5221	2,28	2,28	-	-
5222	1,82	1,31	0,76	1,85
5223	1,73	1,73	-	-
5224	2,23	1,98	1,23	2,74
5225	2,40	2,40	-	-
5226	0,92	0,92	-	-
5227	2,41	1,89	0,99	2,30
5231	1,66	1,66	-	-
5232	1,08	1,08	-	-
5233	3,32	2,03	0,86	3,20
5241	4,47	1,42	0,74	3,04
5242	10,92	2,28	0,91	4,96
5243	7,76	6,27	4,26	8,29
5244	5,03	2,83	0,88	5,75
5245	3,82	3,19	1,86	4,23
5246	6,71	3,39	1,60	7,42
5247	5,28	1,78	0,55	2,77
5248	5,52	3,81	0,99	6,92
5261	3,04	3,04	-	-
Total	*2,74*	*2,87*	*0,55*	*8,30*
3-DIGIT ATECO				
521	16,32	5,20	4,72	5,67
522	2,88	1,97	0,92	2,41
523	2,35	1,73	1,08	3,32
524	7,60	6,19	3,82	10,92
526	3,04	3,04	-	-
Total	*2,74*	*3,62*	*0,92*	*10,92*
15 ISTAT SPECIALISED GROUPS				
1	5,44	1,83	0,76	2,28
2	1,70	1,37	1,08	1,66
3	10,71	2,05	0,74	4,96
4	7,76	6,27	4,26	8,29
5	3,89	1,60	0,88	3,42
6	3,98	3,98	3,98	3,98
7	3,14	2,45	1,42	3,49
8	6,92	6,92	6,92	6,92
9	5,33	4,35	2,95	5,75
10	6,71	3,39	1,60	7,42
11	3,32	2,03	0,86	3,20
12	5,28	1,78	0,55	2,77
13	3,89	3,04	1,86	4,23
14	4,23	3,58	2,90	4,25
15	4,88	3,39	0,99	5,53
Total (3)	*6,77*	*3,20*	*0,55*	*7,42*
8 EUROSTAT GROUPS				
E1	5,67	5,23	4,69	5,76
E2	4,72	2,72	2,41	3,03
E3	2,88	1,97	0,92	2,41
E4	2,35	1,73	1,08	3,32
E5	9,44	7,71	4,47	10,92
E6	5,16	5,19	3,82	6,71
E7	5,50	5,40	5,28	5,52

E8	3,04	3,04	-	-
Total	*2,74*	*4,12*	*0,92*	*10,92*

Source: elaboration on ISTAT data (1997).
(1) Means are calculated considering the CV of the ATECO of lower order in each group.
(2) Missing values relate to classes with no lower order group.
(3) Not specialised stores are not included.

Table 3.4 - % share of enterprises with at least one outlet belonging to an associative form by ATECO categories and classes (1997 sample survey)

ATECO 5-digit	% Share	ATECO 4-digit	% Share	ATECO 5-digit	% Share	ATECO 4-digit	% Share
52112	100,0	5211	13,8	52426	0,0		
52113	18,5			52431	6,8	5243	6,1
52114	10,3			52432	1,9		
52121	3,7	5212	4,5	52441	4,9	5244	5,8
52122	5,1			52442	5,9		
52210	3,7	5221	3,7	52443	7,1		
52221	4,4	5222	4,7	52444	0,0		
52222	8,8			52445	6,0		
52230	9,1	5223	9,1	52451	8,3	5245	8,9
52241	7,3	5224	10,7	52452	12,5		
52242	14,6			52453	5,9		
52250	4,1	5225	4,1	52454	8,5		
52260	0,0	5226	0,0	52455	0,0		
52271	5,1	5227	8,1	52461	4,5	5246	3,7
52272	3,2			52462	3,2		
52273	5,7			52463	0,0		
52274	16,7			52464	1,1		
52310	2,2	5231	2,2	52465	9,5		
52320	10,0	5232	10,0	52466	4,9		
52331	10,4	5233	10,4	52471	27,3	5247	6,8
52332	10,4			52472	2,4		
52411	0,0	5241	3,8	52473	6,3		
52412	4,3			52481	10,8	5248	7,7
52413	0,0			52482	9,2		
52414	14,3			52483	5,7		
52421	3,7	5242	4,8	52484	12,5		
52422	19,2			52485	6,3		
52423	4,2			52486	9,4		
52424	2,9			52489	7,1		
52425	0,0			52610	7,1	5261	7,1

Source: elaboration on ISTAT data (1997).

4. Identification of the most discriminant variables for retail trade turnover

A statistical technique to verify the dependence of retail trade turnover from one or more explicative variables, expressed in continue or in discrete categories, is represented by a particular procedure of discriminant cluster analysis run out by the statistical package SPSS. This analysis considers, given the dependent variable to be explained "turnover per enterprise", the iteration of a splitting hierarchical algorithm based on the Ward optimisation: search of the aggregation which consider the maximisation of the ratio between the explained variance (*between* groups) and the residual variance (*within* groups)[9]. This procedure presents the advantages to not impose a linear

---

[9] The technique is the following: k variables potentially explicative of a dependent variable are given; for the i-th of these k variables, provided by c_i categories, as a first step the units related to a couple of categories characterised by a maximum significance value of Fisher's F test are matched before. By this first step the units will be clustered in (c_i-1) groups. This procedure is replied until the original units are grouped in only two final clusters. This procedure is

binding on the kind of relation between dependent and independent variables, and to discover automatically the stratification about the explicative variable which better explains the variability of the examined character. In other words we are using a self-classifying method, by which it's possible to evaluate the goodness of the actual classifications more or less used in practice.

The proposed application refers to the sample used during 1997 by ISTAT in the monthly retail trade survey, composed by 7.731 enterprises, of which about 1.200 belonging to the "large distribution" (NACE 52.1) and the remaining to the specialised retail trade shops (NACE 52.2, 52.3 and 52.4, so that sales of used goods, sales not in fixed outlets and repairs are excluded).

As we said the dependent variable is constituted by turnover referred to the whole year 1997, and we considered as possible discriminating variables the following eleven: NACE at the level of four digits (class), the special classification used by ISTAT for the monthly retail trade survey based on 17 groups (the 15 specialised mentioned in the previous paragraphs and two additional not specialised sectors, with food and not food products predominating), year of business start-up, geographical region (20 modalities), number of employees, belonging to a voluntary chain (yes/no), a buying group (yes/no), a franchising chain (yes/no), localisation in a trading centre (yes/no), use of extra-time openings (yes/no) or use of sales promotions (yes/no) in at least three months of the year.

Main results are resumed in the table 4.1. The model explains only the 30,3% of variance: it means that all the above mentioned variables are not able, in reality, to identify very homogeneous strata in which enterprises are characterised, more or less, by *the same average level of turnover*. Let's note that this is one of the most common and strongest assumptions, often considered true without any further investigation, that is done in every research context in which a sample survey must be used to observe the domain which we are interested in.

The procedure is based on three steps, on the basis of which further branches are discovered and detailed.

1. At the first step the most important discriminant variable is identified: it's the number of employees, that is positively correlated with the average turnover. Classes are: 1-2, 3, 4-8, >8, so that they are quite similar to those used currently when stratifying a certain population of enterprises (the official stratification based on employees used in Italy for what concerns retail trade is 1-2, 3-5, 6-9, 10-19, >19).

2. At the second step we find out that the NACE classification is relevant just for the second group (enterprises with 3 employees), that clearly represents only a particular subset of all the existing retail trade firms. These enterprises are broken down in four subgroups: the ones with NACE ranging from 52.11 to 52.31 (not specialised, specialised with food and beverages predominating and specialised with pharmaceutical predominating), those with NACE ranging from 52.32 to 52.43 (specialised with orthopaedic appliances, cosmetics, textiles, clothes, shoes, leather articles predominating), those with NACE ranging from 52.44 to 52.47 (specialised with furniture, lighting articles, electrical household appliances, ironmonger's tools, stationery, books, newspapers predominating) and those with NACE ranging from 52.48 to 52.6 (other products and sales outside shops). It's worthwhile to recall that the Short-term Business Statistics Regulation obliges to calculate monthly indexes for eight groups of products, generally speaking quite different in comparison with the above mentioned detected by the algorithm, with a partial exception for what concerns the group characterised by NACE ranging from 52.44 to 52.47.

---

replied for every variable belonging to the k explicate ones. For every level of breakdown the split variable, respect to the couple of chosen categories with a maximum significance F value, is considered. Then in the following breakdown the variable already used for the previous split will be used no more. If for any level no variable overcomes the significance test, the procedure stops without any further aggregations.

Moreover, the first-step group containing enterprises from 4 to 8 employees can be successfully broken down in four subgroups on the basis of the simpler ISTAT classification based on 17 groups of retail activities, and we note that in this classification the specialised food and beverages shops are left alone (code 1), as the not specialised stores (codes 16 and 17), that in the ISTAT nomenclature are not exactly equivalent to not specialised identified by the NACE. Then we have a group containing the enterprises with codes 2, 3 or 4 (specialised with pharmaceutical, textiles, clothes and leather predominating, in other words the "second necessity" goods after food and beverages) and, finally, a residual group with codes ranging from 5 to 15 (the "other products" in a wide definition). At this second step for the remaining enterprises, very small (1 or 2 employees) or rather big (more than 8 employees), the only relevant variable is the location in a trading centre (if yes we have a quite higher average turnover, as it's reasonable).

3. At the third step other additional ten subgroups are identified, of which six are based on the surface of the shop: surface is fundamental to better specify the subgroups identified on the basis of NACE at the second step, the subgroups of small enterprises and big enterprises not localised in a trading centre and the subgroup of enterprises with employees ranging from 4 to 8 and characterised by the ISTAT groups of products ranging from 5 to 15. Other discriminant, but less important variables are the recourse to sales promotion and the year of business start-up.

Now, which conclusions arise from these empirical evidences? We feel that, as a matter of fact, the complexity of the retail trade sector is too high to guarantee a really successful performance for the NACE classification or, more or less, for any other classification based on these concepts:

- the exact identification of the physical place in which sales are realised;
- the idea that a predominating activity really exists for any specialised enterprise;
- the main aim to highlight the "intrinsic characteristics of the product sold" rather than the "type of utilisation of the product" or, in other words, the consumption function the product is designed to satisfy.

5. A short resume and items to be further discussed

In the internal market framework we are facing a development toward new forms of distribution, less defined when compared to the ones considered by NACE (basically founded on the dualism between specialised and not specialised enterprises).

We guess that to analyse a very complex economic sector we need a set of alternative classifications, each aimed to put in evidence a particular behaviour of the population under study, possibly partially overlapping in order to render easier an attempt to reconvert each one into another of higher level. When we decide to modify NACE we must recognise that:

- real changes occurred in the retail trade sector are detected only after some years from their start-up;
- additional years are generally requested in order to guarantee a general agreement of most countries on the new classes to be added or the old ones to be modified;
- then, other time is needed to practically implement the new classification and to reconvert old historical series into the new ones.

This strategy doesn't seem the most useful and reliable at all. In the light of the previous analysis (we could reply it changing the sample, or putting in the discriminant analysis as dependent variable the year to year change of turnover instead of its average level, but we would obtain very similar results) we suggest to deeply reflect in order to propose additional classifications that, without replacing NACE at all (it's straightforward that NACE remains the basic reference point in each EU country) could improve our degree of knowledge on the retail trade sector and help us to adopt logical schemes more fitted with the definitions commonly adopted by retail enterprises.

Further hypotheses are:

1. getting over the existing rule that uses as exclusive discriminant factors two single trading forms, namely specialised and not specialised retailers, trying to better evaluate new mixed types of retailing;
2. re-classifying some of the typologies using more detailed subdivisions only for those 4-digit headings where empirical evidence shows an imbalance in size confirmed by a high heterogeneity of turnover (as it happens in a lot of 5-digit ATECO); EUROSTAT could play a central role in harmonising such subdivisions.

References

ANITORI P., GISMONDI R., RICCARDINI F., TROVATO G. (1997), *"New" and "Old" Services: An Overview for Some Italian Industries*, paper prepared for the "12th Voorburg Group Meeting", Session 4, Copenaghen.

CECCARELLI M.R., GISMONDI R., MIRTO A.P. (1997), *Late Developments in Some Italian Service Surveys: Comparison of Problems and Advantages*, paper prepared for the "12th Voorburg Group Meeting", Session 3, Copenaghen.

EUROSTAT (1998.1), "Breakdown of Turnover by Product", Document discussed during the *Working Group on Trade Statistics*, March 1998, Eurostat, Luxembourg.

EUROSTAT (1998.2), "Particular Problems in Retail Trade and Other Services", Document discussed during the *European Business Trend Working Party*, May 1998, Eurostat, Luxembourg.

EUROSTAT (1998.3), *The EU Regulation on Short Term Indicators*, final version, May 1998, Eurostat, Luxembourg.

GISMONDI R. (1998.1), "Classification Problems in Retail Trade: Theoretical Aspects and Practical Results", *Istat Documents*, 8, Istat, Rome.

GISMONDI R. (1998.2), "The Impact of the Short-term Business Statistics Regulation", paper to be presented at the *13th Voorburg Group Meeting*, 21-24 September 1998, Rome.

GISMONDI R., MIRTO A.P. (1998), "Classifications and Heterogeneity of Data in Tourism Statistics: the Italian Case", paper presented at the *Joint IASS/IAOS Conference: Statistics for Economic and Social Development*, Aguascalientes, Mexico, September 1998.

ISTAT (1991), *Classification of Economic Activities*, "Methods and Rules", 11, Istat, Rome.

ISTAT (1998.1), *The Retail Trade Monthly Indexes in 1996*, "Information", 56, Istat, Rome.

ISTAT (1998.2), *The New Survey on Retail Trade Sales: Methodological and Innovative Aspects*, "Methods and Rules", 3, Istat, Rome.

ISTAT (1998.3), *How Trade is Changing: Structural Modifications and Employment Trends*, "Topics", 13, Istat, Rome.