



AUTORITÀ GARANTE
DELLA CONCORRENZA E DEL MERCATO

Mauro La Noce, Elisabetta
Allegra, Valerio Ruocco, Federico
M. Capo

**Merger control in Italy 1995-
2003: a statistical study of the
enforcement practice by mining
the text of Authority resolutions**

Temi e Problemi

volume 13
giugno 2006

La serie “Temi e Problemi” ospita contributi prodotti all’interno dell’Autorità Garante della Concorrenza e del Mercato, allo scopo di rendere disponibili a studiosi e alle istituzioni interessate i risultati delle attività di ricerca in corso presso l’Autorità e stimolare la discussione su argomenti connessi alle tematiche istituzionali. I lavori pubblicati nella collana riflettono esclusivamente le opinioni degli Autori e non impegnano la responsabilità dell’Autorità. Commenti sui singoli contributi possono essere indirizzati direttamente agli Autori, all’indirizzo indicato.

The “Temi e Problemi” working papers are preliminary materials prepared within the Autorità Garante della Concorrenza e del Mercato (the Italian Competition Authority), circulated with the aim of stimulating discussion and critical comment. The analyses and conclusions are those of the Authors and do not involve the responsibility of the Autorità. Comments may be directly addressed to the contact Author(s).

Redazione: Pierluigi Sabbatini (coordinatore), Lapo Berti, Renato Ferrandi, Giuseppe Galasso, Giuseppina Mangione, Alessandro Noce, Andrea Pezzoli, Giovanna Ragno.

Original draft: September 2005

This draft: June 2006

Mauro La Noce, Elisabetta Allegra, Valerio Ruocco, Federico M. Capo

Contact author:

Mauro La Noce

Autorità Garante della Concorrenza e del Mercato,

piazza G. Verdi, 6/a, 00198 ROMA - Italy

tel. (39 +) 06 85821 319

fax: (39 +) 06 85452 319

Email: mauro.lanocce@agcm.it

Acknowledgments

This is a preprint of an article submitted for consideration in the International Journal of the Economics of Business © 2006 International Journal of the Economics of Business; International Journal of the Economics of Business is available online at <http://journalonline.tandf.co.uk/>. The authors thank Giulia Antenucci, Sergio Bolasco, Malcolm B. Coate, Fabio Massimo Esposito, Mario Forni, Roberto Franzosi, Alberto Heimler, Pierluigi Sabbatini and Francesco Silva for helpful comments provided on an earlier draft; Simonetta Schettini and Manuela Villani for their invaluable assistance on data collection; Francesco Baiocchi and Alessio Canzonetti for providing excellent programming support. The usual disclaimer applies: the conclusions expressed in this paper are exclusively those of the authors, as well as any possible mistakes. The views expressed here do not necessarily represent those of the Autorità Garante della Concorrenza e del Mercato (the Italian Competition Authority).

Suggested citation:

La Noce, M., Allegra, E., Ruocco, V. and Capo, F.M. (2006), "Merger control in Italy 1995-2003: a statistical study of the enforcement practice by mining the text of Authority resolutions", Autorità Garante della Concorrenza e del Mercato, Temi e Problemi, 13.
Available at: <http://www.agcm.it/temieproblemi.htm>.

Mauro La Noce(*), Elisabetta Allegra(*), Valerio Ruocco(*), Federico M. Capo(**)

Merger control in Italy 1995-2003: a statistical study of the enforcement practice by mining the text of Authority resolutions

Abstract

The paper provides a consistency analysis of the decisions taken by the Italian Competition Authority within the 1995-2003 period, concerning its control over mergers and acquisitions. An ample database was constructed using the tools of textual analysis. This data base was generated in a semi-automatic manner and handled with statistical methods, the final aim being the econometric specification – through a *logit* model – and estimation of the “de facto” decision model of the Authority, thereby making its principles more transparent and predictable. Model simulation exercises, dependant on multivariate market scenarios, helped estimate the threshold levels triggering Authority intervention.

JEL Classification: K21, L40, D78, C25, C81

Keywords: Antitrust, Competition Law, Merger, Merger control, Text-mining

Sintesi

Il lavoro si propone di verificare la coerenza delle decisioni assunte dall’Autorità *antitrust* italiana nel periodo 1995-2003 in materia di controllo delle fusioni e acquisizioni fra imprese. Il ricorso a metodi e strumenti innovativi di analisi del testo ha reso possibile la ricostruzione di un’ampia base dati, generata in maniera semi-automatica. La successiva specificazione e stima econometrica di un modello *logit* delle decisioni ha permesso di valutare la rilevanza dei principali fattori che ne determinano gli esiti, contribuendo a una migliore conoscenza degli orientamenti seguiti dall’Autorità in questo campo. Esperimenti di simulazione del modello, basati su specifici scenari di struttura dei mercati, hanno quindi consentito di determinare soglie di attenzione degli stessi indicatori di struttura oltre le quali l’intervento dell’Autorità diviene un evento molto probabile.

(*) Autorità Garante della Concorrenza e del Mercato.

(**) Intern at the Autorità Garante della Concorrenza e del Mercato (September 2004 – March 2005).

1. Introduction

The Italian Competition Authority's power regarding merger control are spelled out in the 10 October 1990 law no. 287 (the Competition and Fair Trading Act). Its enforcement has resulted in a *corpus* of decisions concerning concentrations between undertakings, numbering 5,500 as of October 2004. A retrospective analysis of such data, embodied in case resolutions, provides specific market frameworks within which the Authority's intervention to stem undesired concentration effects on the relevant markets becomes a predictable event. As recently posited, "*unpredictable merger policy can have real costs in terms of increased unemployment and decreased investment levels*"¹, and transparency of procedures and criteria used by competition agencies is instrumental in achieving a higher level of law-enforcement predictability.

This paper investigates the determinants of the Authority's merger control decisions, through an econometric analysis of a high volume of information regarding the Authority's own assessment of the competitive consequences of corporate concentrations. The information originates from data systematically gathered by the Authority's case files *and* from text-mining techniques² applied to the full text of the adopted resolutions.

The use of these techniques in such a context is new inasmuch as, through consistent and verifiable text queries, it extracts from "unstructured" language – namely, the full text of Authority resolutions – the information of interest, enabling the coding of textual data and the subsequent use of statistical techniques to quantitatively appraise the Authority's decision model.

The subjectivity of the resulting classification of the relevant information has also correspondingly been reduced, as well as (not less importantly) the number of qualified human "code makers". This has permitted the coding of a high volume of documents, thanks to the extensive use of statistical-linguistic resources (standard language lists, electronic dictionaries, specific grammars, etc.) and through a pre-treatment of the texts (normalization, segment identification, pre-tagging of grammatical locutions, etc.)³.

¹ Voigt and Schmidt (2005), p. 5, with reference to the European merger policy.

² The nascent field of text-mining techniques has been developing since mid-'90s, in order to deal with information overload from large text collections and its increasingly availability in electronic format, by linking the *information retrieval* and *information extraction* processes, preceded by the computer analysis and manipulation of natural language texts. Bolasco (2004), p. 15. "[T]he text-mining process is preceded by two essential phases: (1) The pre-processing phase, where text retrieval, formatting and filing is done. (2) The lexical processing involves identification and lemmatisation of words. Following these two phases, the actual text-mining processing is extremely diversified, as it is strictly linked to the objectives to be achieved. They are:

automatic analysis of documents and their categorisation/classification for successive information retrieval;

search for relevant entities for information extraction;

formulation of queries in natural language, interpreted by NLP [Natural Language Processing] processes based on algorithms of artificial intelligence;

processing of multi-lingual text for the retrieval of information independent of the original language of the documents."

Bolasco et al. (2004c).

³ See Baiocchi et al. (2005) for a non-technical description of the text-mining application on the textual collection ("document warehouse") of the Authority's decisions.

The estimation of the parameters of the decision model, and its simulation exercises based on multivariate market scenarios, have confirmed the relevance of traditional market structure indicators, such as post-merger market shares of the merging parties, market concentration levels and their variation over time. It has also been posited that the threshold levels triggering Authority intervention are determined by the concurrence of other important structural characteristics of the markets affected by the merger, such as the presence of barriers to entry, occurrence of a countervailing buyer power, ease of collusion, as well as by possible vertical and scope effects of the merger. Finally, the relevance of dynamic factors influencing the Authority's resolutions has also been inferred, mainly with reference to how past decisions may influence the subsequent ones.

This paper consists of four sections and two appendices organized as follows. In the first section we review the empirical literature on antitrust control of mergers and acquisitions taken from a number of jurisdictions throughout the world. The second section deals with a brief description of the data extracted from the Italian Competition Authority's resolutions during the 1995-2003 period. Then, in the third section the decision model of the Authority is specified and econometrically estimated. Finally, in the fourth section we evaluate the threshold levels triggering Authority intervention. The content and structure of the database, as well as a non-technical overview of the text-mining procedures used in this study, are contained in the two appendices.

2. Empirical literature

The empirical econometric evidence on what determines antitrust decisions is quite scant. The relevant information is stored in the competition agencies' archives and is seldom available.

Few papers, mostly from the '90s, have analysed the North American experiences – namely US and Canada⁴ – as well as those of New Zealand and the United Kingdom. More recently, econometric studies of the EU Commission decisions regarding merger control⁵ have appeared, and interesting initiatives have been undertaken by the US and UK antitrust authorities aiming to assess the consistency and consequences of their merger control enforcement activity.

These analyses and their results largely reflect the specific jurisdictional norms and the criteria implementing them, in a few cases taking the form of guidelines. An example is provided by the United States, where decisions were traditionally oriented at assessing concentration effects in terms of collusion (coordinated interaction): the firms remaining after concentration tend to share *collective* rather than a *unilateral* market power, due to the very changes in the

⁴ See bibliography, ordered by jurisdiction.

⁵ See also a few quantitative studies analysing the *Bundeskartellamt* (Germany) and the *Office of Fair Trading* (United Kingdom) decisions, but related to cartels and abuses of a dominant position (no to merger control enforcement). Lauk (2002), for Germany; Davies et al. (1999), for the United Kingdom. With reference to the European Commission's non-merger decisions, see the work (still in progress) by M.P. Schinkel, M. Carree and A. Günter, "European antitrust policy: an empirical analysis of Commission decisions and their appeal histories 1964-2001", available at <http://www.diw.de/english/produkte/veranstaltungen/earie2004/papers/docs/2004-462-V01.pdf> (preliminary version for EARIE 2004, Berlin).

market structure. In the early 1990's, the revision of the US merger guidelines brought forward the relevance of unilateral effects of mergers in markets with differentiated products. Following the initiative promoted by the *U.S. Federal Trade Commission (FTC)* on "Transparency and the Horizontal Merger Review Process", tabulations on enforcement data spanning the 1996-2003 period have been released, by thorough analysis of FTC staff memoranda.⁶ Quite a large number of studies have recently appeared, modelling, with econometric techniques, the FTC decisions (Kwoka, 2004, Coate and Ulrick, 2005, Coate, 2005a, b, c). The two theories of competitive concern – coordinated interaction and unilateral effects – have been tested (Coate, 2005b), with the market concentration index (HHI)⁷ and the number of significant rivals in the market representing proxies, respectively, of ease of collusion and unilateral effects. Simplifying the evaluation approach, "[h]igh concentration and merging firms with nontrivial shares constitute legitimate grounds for concern over coordination"⁸; on the contrary, within the hypothesis of unilateral effects "concentration plays no role for the simple reason that cooperative action among firms is not an issue"⁹.

The likelihood of enforcement action has been derived from the parameters of the estimated models. Kwoka (2004) implies that the probability of US antitrust intervention increases by 65.5% with an increase of 600 units of the HHI concentration indicator (starting at 2,400). Coate and Ulrick (2005) and Coate (2005b) depict a more complex scenario with more explanatory variables, such as the number of significant competitors¹⁰, barriers to entry, customer complaints, existence of "hot" evidence¹¹ and of anticompetitive event studies¹². As an example of simple model prediction of FTC enforcement, "a collusion case with a post-merger HHI of 3,611 has a 50 per cent chance of enforcement, while a deal affecting an industry with 3 pre-merger competitors would face an enforcement probability of 38 per cent", entry barriers being present¹³. This same model does not provide any significant evidence of

⁶ This review examines 151 horizontal transactions, which encompass 784 relevant markets. For a description of the data, see Federal Trade Commission (2004). The review only deals with *second request* investigations, i.e., a small number of transactions where the agency requests additional documents and information – namely, the "second request" – to supplement the parties' initial filings. As a result, this sample of cases is not representative of the entire U.S. merger control enforcement activity, but only refers to the subset of cases where anti-competitive concerns have arisen.

⁷ Market concentration is usually measured (in antitrust analyses and enforcement) by the Herfindahl-Hirschman Index (HHI), calculated by summing the squares of the individual market shares of all the participants (e.g., U.S. Department of Justice/Federal Trade Commission, 1992 Horizontal Merger Guidelines).

⁸ Kwoka (2004).

⁹ More specifically, the role of demand substitutability in explaining unilateral effects is highlighted, in its turn "a function of own and cross demand elasticities. Since these are rarely available, the diversion rate – a measure of fraction of sales likely to go to the alternative product – can be informative. And it should be noted that under certain assumptions, market shares help in estimating the magnitude of the diversion ratio. [...] [U]nilateral effects depend upon elasticities and diversion factors, which are best partially informed by market shares but otherwise not closely related to traditional structural characteristics". *Id.*, pp. 4 and 6.

¹⁰ "The operative definition of a significant competitor was a firm whose independence could affect the ability of the merged firms to achieve anticompetitive outcome." When dealing with coordinated effect hypotheses, "a significant competitor would be noted as a required participant in the collusive group"; in cases of unilateral effects, "a significant competitor would be identified as a close rival to the merging parties". Coate and Ulrick (2005), p. 5.

¹¹ "Hot" evidence is represented by "a document, submitted by one of the merging parties, projecting that the merger would result in an adverse price or non-price effect on competition in the relevant market". *Id.*, p. 6.

¹² "Event studies focus on some structural or strategic change (i.e., natural experiment) in the market (or a related market) that approximates the competitive effect of the merger". Coate (2005b), p. 11.

¹³ Coate (2005b), p. 29.

the effect of merger-related efficiencies¹⁴, as well as any clear influence of econometric analysis to support the FTC evaluation of the competitive effects¹⁵.

The Canadian Antitrust (Khemani and Shapiro, 1993, 1994) seldom takes market concentration indicators as the basis of its decisions. Hence it considers – on a case-by-case basis – the market shares of the participants in the merger as a measure of market structure. As for other variables explaining decisions, barriers to entry and the degree of openness to external trade play an important role.

As for the United Kingdom, the *Fair Trading Act* of 1973 dictates that, in order to allow an operation, merger controls should establish that it is not contrary to the “public interest”¹⁶. A study based on a sample of 70 decisions, taken by the *Monopolies and Mergers Commission* (MMC) between 1974 and 1990, reaches the conclusion that market shares do not, by themselves, bear influence upon decisions (Weir, 1992). The latter appear to be determined by other factors, such as the MMC disposition to agree on the parties’ arguments, the hostile nature of acquisitions, the deemed absence of effects on competition as well as on prices.

It is worth mentioning, even though not in the econometric field, a recent report commissioned by the UK Antitrust Authority on ex-post evaluation of mergers, carried out through the retrospective analysis of a small sample of transactions referred to the Competition Commission and cleared without remedies within the 1991-2002 period¹⁷. From the case files, the following factors have been cited in clearing mergers (ordered by decreasing number of citations): inter-firm rivalry, ease of entry/expansion, buyer power, failing firm/weak competitor, substitution from outside the market, regulation.

In New Zealand, 200 decisions by the *Commerce Commission* were examined over the 1991-1996 period (Strong et al., 2000). Two factors seem to play a role: the combined market shares held by the two sides and the presence of barriers to entry. More specifically, in the presence of barriers to entry, a 70% total of market shares is most likely to bring an unfavourable decision.

As for the European Union, two studies recently appeared regarding the Commission’s merger control enforcement decisions, based on an econometric approach using binomial *logit* specifications¹⁸. The first study uses for its estimation a sample of 96 cases over the 1990–

¹⁴ However, the author notes that “the impact of efficiencies has [possibly] been built into the model through [the] customer complaint [variable]”. *Id.*, p. 32.

¹⁵ According to the author, “structural econometrics methods do not seem to affect the outcome of the investigation. While statistical insignificance does not prove statistical studies never affect decision makers, the lack of a clear effect suggests that parties should not waste their resources on empirical analyses when good industry data are not available. [...] [T]he FTC clearly recognizes the answers to the real-world questions depends on a myriad of institutional facts related to the competitive organization of the market”. *Id.*, pp. 35-36.

¹⁶ In June 2003 the *Enterprise Act* entered into force, which substantially reforms merger control enforcement process in the United Kingdom, more in line with EC rules, and the *Office of Fair Trading* was established as an independent statutory body.

¹⁷ Office of Fair Trading (2005).

¹⁸ Lindsay et al. (2003), and Bergman et al. (2003). Other interesting econometric analyses of the determinants of the European Commission’s merger control decisions have followed quite a different approach, relating those decisions to the stock market reactions to the merger, and not strictly relying on the Commission’s own analysis to assess competitive factors (Neven and Röller, 2002; Duso et al., 2003; Aktas et al., 2004). These studies are not reviewed here, and we refer the reader to other papers commenting on them (e.g., Office of Fair Trading, 2005, p. 106; Bergman et al., 2003, p. 6). Finally, quite a large number of non-econometric studies have been realized on the

September 2002 period, and the second relies on 245 cases over the January 2000–June 2002 period. Both studies point to the statistical relevance of total market shares of the merging parties, as well as of barriers to entry, in influencing the Commission’s decisions. One study (Lindsay et al., 2003) establishes low significance to buyer power – as a countervailing effect with respect to the market power exercised by the parties to the concentration – and to market concentration, as proxied by the HHI indicator. Both findings are attributed to the low quality of data. The other study (Bergman et al., 2003) points to a higher probability of denial as a consequence of collusive behaviour triggered by the merger (cases of collective dominance examined by the Commission). The market concentration indicator is seldom a significant proxy of the expected collusive behaviour. Ease of collusion is revealed to be significant and with the expected sign only when explicitly mentioned in the Commission’s decisions, and directly considered as an explanatory variable in the econometric model.

Some common results emerge from the reviewed studies. First, strong evidence has been provided for the role of entry barriers, whose absence significantly affect the explanatory power of market shares and concentration indexes. Secondly, such findings seem to reflect to some extent a low reliability of the available information and inaccuracy in the data, mainly when proxy variables are used. The most significant results emerge from studies where a strict “consistency approach” (of the competition agencies’ decisions) applies and *“data used to determine whether a merger is likely to be cleared or blocked comes from the competition authorities’ own reports”*¹⁹.

This work tackles these specification issues by using data on market structure directly originating from the Italian Competition Authority’s merger control enforcement activity (definition of the relevant market, market shares of the merging parties and their competitors), as well as by cross-examining the merger evaluation criteria by means of text-analytic techniques applied to the entire “document warehouse” of the Authority’s resolutions.

3. Overview of the data

The criteria for the competitive assessment of merger cases adopted by the Italian Competition Authority do not diverge considerably from those applied by other jurisdictions, as sketchily described in the overview of the empirical literature. The post-merger market shares of the merging parties and a market concentration index (HHI or others) are approximate indicators of, respectively, unilateral and coordinated effects of the transaction. Other relevant factors can be inferred from the analysis of the Authority’s resolutions related to the (few) formally

subject matter; see, among all, Neven et al. (1998), and, more recently, Verouden (2004) and Voigt and Schmidt (2005).

¹⁹ Office of Fair Trading (2005), p. 105. This approach is criticized on the grounds that the data used is correlated with the competition authorities’ decisions. We do not share this view, since such a correlation does not imply any bias in the estimated model. Only correlation between explanatory variables has to be avoided in order to obtain unbiased estimates of the model’s parameters.

investigated cases. Thirty-four cases were considered within the 1995-2003 period, related to decisions taken after that a formal investigation had been commenced, and a number of relevant elements have been detected, such as barriers to entry, countervailing buyer power, probability of collusive behaviour, horizontal and vertical integration resulting from concentration.

Defining the unit of analysis has forced us to focus on the mechanics of the concentration effect on markets. Indeed, the final decision bears on the “case” as a whole. Yet the evaluation must deal with the effects of concentrations on several markets. The elementary units to be analysed are, therefore, the markets affected, where each relevant market is identified through both its product and geographic characteristics. The product dimension is defined by the ISTAT (the Italian National Institute of Statistics) economic activity code (ATECO), with an extension when needed, and, in some cases, further specified by its primary distribution channel. The geographic dimension is represented by the code of the administrative territorial type and its specific area code. As such, the individual observation (the elementary unit of analysis) used in the decision model is the following sequence of attributes: *case reference – product code – administrative geographic unit – area code – distribution channel*.

The model response is a binary variable, where 0 represents the Authority’s approval, 1 the operation prohibition or its approval with commitments by the concerned undertakings. The latter may well be referred to a subset of the markets involved. For each case, the 1 response is, therefore, associated with markets where the Authority’s “intervention” is needed in order to prevent possible anti-competitive effects of the examined merger.

The Authority’s decision model can thus be approximated by its logistic transformation, the dependent variable being the likelihood of Authority intervention, and as such constrained within the 0-1 range of values²⁰.

3.1 The model data set

The Authority has assembled a relational database named GATC (an acronym for “Gestione delle Attività di Tutela della Concorrenza” – Competition Activity Management) depicting the entirety of cases examined since 1990, when the Authority was founded. Such a database is routinely updated (see Appendix I for an explanation of its content).

A number of explanatory variables were directly taken from this database:

- The distribution of market shares of the parties to the concentration and their main competitors allow for the calculation of the following quantitative variables:

²⁰ As to the mathematical specification and the econometric estimation method, the reader may refer to Maddala (2001), pp. 322ff, and to Allison (1999) for a more specific and practical approach.

- the post-merger combined market share of the merging parties: it is calculated as the sum of their pre-merger market shares, and gives a first evidence of their market power;
 - the concentration levels: indexes from CR1 to CR5²¹ may be calculated, while the HHI index is much more demanding in terms of data;
 - variation of the concentration levels: the increase in concentration as measured by the HHI index can be easily calculated by doubling the product of the market shares of the merging firms, and it represents their overlapping on the affected markets. When equal to 0, any horizontal effect of the concentration is lacking.
- Geographic dimension of the market: two sets of *dummy* variables, representing, respectively, the sub-national the supra-national dimension of the affected markets. We assume that the wider the geographic market, the lower the probability of anti-competitive effects of the concentration.
 - Distribution markets: most of the cases involving wholesale or retail trade markets are related to the acquisition of small mercantile establishments, with no competitive effects. A large number of cases was notified to the Italian Authority, since the notification thresholds refer to the value of turnover rather than (as it would appear more correct) to trade margins.

While the GATC database focuses on the definition of relevant markets, and on quantitative data related to the structure of the markets, it does not contain qualitative information regarding other important factors influencing Authority decisions.

In order to locate and classify this type of information, the entire documentary archive containing the full text of decisions on non-investigated cases has been examined with a statistical program of text classification: TALTAC²². This tool, appropriately modified to suit our specific data context, established a logical operators set enabling us to build query procedures which permit the retrieval of qualitative information “buried” in the decisions’ text²³. Due to the experimental nature of the procedure, only those decisions with a relatively homogeneous textual structure were considered; the 1995 year was chosen as the starting point because it is the first year for which such homogeneity applies. A set of 3,646 final decisions (regarding 9,473 relevant markets, since one merger could affect a number of different markets) was examined with TALTAC.

²¹ CR_n is the percentage of total market shares accounted for by the leading *n* firms.

²² “TALTAC stands for Automatic Lexico-Textual Treatment for Content Analysis. TALTAC has been realised by Sergio Bolasco, Francesco Balocchi and Adolfo Morrone at the University of Rome “La Sapienza”. TALTAC is a program for statistical and lexical analysis of textual data that allows treatment of a linguistic corpus according to a lexicometric criteria, aimed at the content analysis of text. TALTAC is useful for elaborating free responses to open questions in sample surveys, non-directive interviews and focus groups; collections of articles of press surveys; messages and information on internet; political discourse or advertising and media language. It is also useful for exploring the annual reports of public institutions (UN, EC, Bank of Italy etc.), bibliographic, documentary, literary or essay collections of the most different kind.” See <http://www.taltac.it/en/index.html>. See also Bolasco (1999), Chapter 7 (Analysis of textual data).

²³ Appendix II contains a non-technical description of the programming tools and analytical procedures used in this paper, as well as the list of the queries set in order to extract the information of interest from of the Authority’s decisions.

In this fashion we have identified 9,955 units of analysis out of 3,680 cases²⁴, thus an average of 2.7 units per case (Table 1). We have 3,035 excluded observations where the market shares of the merging parties were missing. When only one of the two parties' share was indicated in the database, we have assumed the other one as scarcely relevant and given it the value of 0 (2,169 observations). As a result, the model data set contains 6,920 observations, within the 1995-2003 period.

TABLE 1

ITALIAN COMPETITION AUTHORITY'S MERGER ENFORCEMENT DECISIONS AND THE MODEL DATA SET 1995-2003

	No. of cases (a)	No. of observations (b)	No. of obs. / no. of cases
Authority decisions:			
- total	3,680	9,955	2.7
- non-investigated mergers	3,646	9,473	2.7
- formally investigated mergers (c)	34	482	14.2
Model data set:			
- total	2,764	6,920	2.5
- Authority approvals	2,736	6,802	2.5
- Authority interventions (c)	28	118	4.2

Source: Italian Competition Authority.

Notes:

(a) Merger cases examined by the Authority within the period. Cases withdrawn on the companies' request were excluded, since only incomplete information is available.

(b) Elementary units of analysis: *case reference – product code – administrative geographic unit – area code – distribution channel*.

(c) The totals of formally investigated cases (34) and Authority intervention cases (28) do not match, since six formally investigated cases were authorized at the conclusion of the investigation.

The use of the lexical-textual analysis tool (TALTAC) permitted us to consider a number of important qualitative variables influencing the Authority's decisions²⁵. Firstly defined at the case level, these variables have been further detailed, when significant, up to the level of the elementary unit of analysis mentioned above:

- Barriers to entry: *dummy* variable set at 1 when the text of the Authority resolution contains expressions which represent factors that prevent firms from entering the relevant market (otherwise 0).
- Countervailing buyer power: a *dummy* variable equal to 1 when the wording of the Authority's resolution indicates the existence of competitive pressure on the merging parties exercised by their customers (otherwise 0).

²⁴ Corresponding to 3,646 non-investigated cases and 34 formally investigated cases. The content of the 34 Authority's resolutions was manually classified, since their textual structure was quite complex and not homogeneous with respect to the cases dealt with by using TALTAC.

²⁵ Figure 4 in Appendix II reports the full list of queries used to extract the information of interest and classify the documents.

- Vertical and horizontal (scope) effects: this *dummy* variable has been set at 1 both when vertical effects of the concentration have been identified in the text of the Authority's resolution, and when the decision revealed the presence of competitors already integrated into upstream or downstream markets (otherwise 0). Vertical effects have been assimilated to scope effects only in those cases where a formal investigation had been concluded²⁶.
- Coordinated effects: *dummy* variable set at 1 when the concentration increases the likelihood of collective dominance, i.e., the firms are able to coordinate their behaviour. Evidence has been analysed and found only after the case has formally been investigated.

3.2 Univariate data analysis

In the period 1995-2003 the data set includes 6,920 observations, related to 2,764 merger cases. The Authority's intervention has concerned 118 out of the total 6,920 observations (1.71%). The temporal distribution of the total number of cases (Table 2) shows some peaks in 1997, 1999 and 2002, mostly depending on formal investigations on concentrations affecting relatively large numbers of geographic markets with a local dimension (ready-mixed concrete, banking, fresh milk, motorway catering markets). The temporal distribution of merger cases also gives evidence of a considerable rising trend from 1995 to 2001, afterwards declining.

TABLE 2
NUMBER OF MERGER CASES AND NUMBER OF OBSERVATIONS BY AUTHORITY'S FINAL DECISIONS
(INTERVENTION/NON-INTERVENTION)

(No. of transactions)

Authority intervention	1995	1996	1997	1998	1999	2000	2001	2002	2003	Total
Yes	-	3	6	2	2	4	4	5	2	28
No	146	174	176	189	310	385	484	431	441	2,736
Total	146	177	182	191	312	389	488	436	443	2,764

(No. of observations)

Authority intervention	1995	1996	1997	1998	1999	2000	2001	2002	2003	Total
Yes	-	3	25	10	23	9	8	38	2	118
No	333	479	505	583	708	883	1,219	1,040	1,052	6,802

²⁶ The use of this variable in the model, as well as its effect on the response variable, requires further clarification. Where a formal investigation has not been commenced, the possible occurrence of vertical effects of the merger (automatically classified) is often offset by competitors vertically integrated into upstream or downstream markets. In these cases, anti-competitive consequences of the concentration are not expected. The retrieval of this concept from the text of the Authority's resolutions simply indicates that these effects have played an important role in determining the final decision (it would not have been mentioned otherwise). Instead, when a formal investigation has been concluded, vertical effects of the merger (manually classified) – as well as horizontal effects extending the scope of product supply – will strengthen the market power of the merging firms, and this effect is not deemed to be significantly offset by vertically integrated and/or diversified competitors. With reference to the Authority's decision model, both effects act in the same direction, i.e., it is more likely that the Authority will challenge the merger.

Total	333	482	530	593	731	892	1,227	1,078	1,054	6,920
(% of no. of observations)										
Authority intervention	1995	1996	1997	1998	1999	2000	2001	2002	2003	Total
Yes	-	0.62%	4.72%	1.69%	3.15%	1.01%	0.65%	3.53%	0.19%	1.71%
No	100%	99.38%	95.28%	98.31%	96.85%	98.99%	99.35%	96.47%	99.81%	98.29%
Total	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%

Figures in Tables 3 to 10 represent a first assessment of the importance of each of the factors depicted in the previous paragraphs in determining the Authority's final decisions.

Barriers to entry (Table 3) play a decisive role in orienting antitrust decisions. By interpreting frequencies in terms of probability, the likelihood of Authority intervention is as high as 55.24% (79 cases out of 143) when factors that prevent firms entering the relevant markets have been detected, otherwise it is highly unlikely (0.58%, 39 markets out of 6,777)²⁷.

Only in two cases has **buyer power** not been able to countervail the market power of the merging parties, calling forth the Authority's intervention (Table 4)²⁸. As expected, the likelihood of intervention is lower when there is evidence of buyer power (0.91%) than otherwise (1.73%).

As much as 10% of markets have required the Authority's intervention when evidence has been found of **vertical or horizontal (scope) effects** as a consequence of the merger (Table 5). Where such effects are absent, the likelihood of intervention is significantly lower (1%).

Very few cases (5 out of 6,920) have been associated with hypotheses of **coordinated effects** (Table 6). This factor, however, acts in the expected direction, since in 4 cases out of 5 the merger has been challenged.

As for the **geographic dimension** of the relevant markets, expectations are confirmed. Wider markets correspond to a lower likelihood of intervention (Table 7): from 0.19% for supra-national markets to 2.04% for sub-national ones.

Mergers affecting **distribution markets** are rarely challenged: the likelihood of Authority intervention is 0.49%, as compared with 2.47% for non-distribution markets (Table 8).

There is evidence to a positive relationship between **post-merger market shares** of the merging parties and the likelihood of the Authority's intervention (Table 9): from 0.50% in the 0-40% range of market shares, to almost 11% in the 40-70% range, up to 25% in the highest rank (70-100%).

In a similar manner, higher **increases of concentration levels** are associated with a higher percentage of mergers challenged (Table 10). The likelihood of Authority intervention rises to 40% when the variation of the HHI index increases in value up to 1,000.

²⁷ With reference to the 39 cases challenged by the Authority even though barriers to entry did not occur, other factors played a decisive role in the final decision: the absence of a significant countervailing buyer power, the occurrence of vertical and/or scope effects (50% of the total), a post-merger combined market share always greater than 48%.

²⁸ In these two cases, the occurrence of barriers to entry and significant vertical effects was ascertained.

TABLES 3-10

UNIVARIATE ANALYSIS OF THE AUTHORITY'S DECISION CRITERIA 1995-2003 (NO. OF OBSERVATIONS)

TABLE 3

ENTRY BARRIERS (PRESENCE/ABSENCE)

Authority intervention	absence	presence	total
Yes	39	79	118
No	6,738	64	6,802
Total	6,777	143	6,920
%			
Yes	0.58%	55.24%	1.71%
No	99.42%	44.76%	98.29%
Total	100%	100%	100,00%

TABLE 4

BUYER POWER (PRESENCE/ABSENCE)

Authority intervention	absence	presence	total
Yes	116	2	118
No	6,585	217	6,802
Total	6,701	219	6,920
%			
Yes	1.73%	0.91%	1.71%
No	98.27%	99.09%	98.29%
Total	100.00%	100.00%	100.00%

TABLE 5

VERTICAL AND SCOPE EFFECTS (PRESENCE/ABSENCE)

Authority intervention	absence	presence	total
Yes	68	50	118
No	6,370	432	6,802
Total	6,438	482	6,920
%			
Yes	1.06%	10.37%	1.71%
No	98.94%	89.63%	98.29%
Total	100.00%	100.00%	100.00%

TABLE 6

UNILATERAL/COORDINATED EFFECTS

Authority intervention	unilateral effects	coordinated effects	total
Yes	114	4	118
No	6,801	1	6,802
Total	6,915	5	6,920
%			
Yes	1.65%	80.00%	1.71%
No	98.35%	20.00%	98.29%
Total	100.00%	100.00%	100.00%

TABLE 7

GEOGRAPHIC MARKET SIZE

Authority intervention	sub-national	national	supra-national	total
Yes	72	44	2	118
No	3,453	2,285	1,064	6,802
Total	3,525	2,329	1,066	6,920
%				
Yes	2.04%	1.89%	0.19%	1.71%
No	97.96%	98.11%	99.81%	98.29%
Total	100%	100%	100%	100%

TABLE 8

DISTRIBUTION MARKETS

Authority intervention	distribution markets	other markets	total
Yes	105	13	118
No	4,151	2,651	6,802
Total	4,256	2,664	6,920
%			
Yes	2.47%	0.49%	1.71%
No	97.53%	99.51%	98.29%
Total	100 %	100%	100%

TABLE 9

POST-MERGER MARKET SHARES OF MERGING PARTIES

Authority intervention	0-40%	40-70%	70-100%	total
Yes	32	40	46	118
No	6,335	327	140	6,802
Total	6,367	367	186	6,920
%				
Yes	0.50%	10.90%	24.73%	1.71%
No	99.50%	89.10%	75.27%	98.29%
Total	100 %	100%	100%	100%

TABLE 10
CHANGE IN HHI

Change in HHI	Authority intervention		Total
	Yes	No	
0	27	3,508	3,535
0-100	2	2,546	2,548
100-500	30	605	635
500-1000	20	103	123
1000-2000	20	27	47
≥ 2000	19	13	32
Total	118	6,802	6,920
%			
0	0.76%	99.24%	100.00%
0-100	0.08%	99.92%	100.00%
100-500	4.72%	95.28%	100.00%
500-1000	16.26%	83.74%	100.00%
1000-2000	42.55%	57.45%	100.00%
≥ 2000	59.38%	40.63%	100.00%
Total	1.71%	98.29%	100.00%

All these findings can be represented in terms of *odds*²⁹ and *odds ratios*³⁰, i.e., the probability of a specified outcome with respect to the adverse probability. The *odds ratios* – as measures of the relationships of pairs of dichotomous variables (e.g., presence/absence, national/sub-national, etc.) – of the so-far commented variables are shown in Table 11.

TABLE 11
LIKELIHOOD (ODDS RATIO) OF AUTHORITY INTERVENTION WITH RESPECT TO NON-INTERVENTION DECISIONS
1995-2003

Characteristics of market structure	Odds ratio
Entry barriers (presence of)	211.5
Buyer power (presence of)	0.52
Scope and vertical effects (presence of)	10.8
Coordinated effects (presence of)	238.4
Geographic market size (w.r.t. the national dimension):	
sub-national	1.08
supra-national	0.10
Distribution markets	0.19
Post-merger market shares of merging parties (w.r.t. the 0-40% range):	
40-70%	24.3
70-100%	65.4
Change in HHI (w.r.t. the value of 0):	
0-100	0.10
100-500	6.4
500-1000	25.2
1000-2000	96.2
≥ 2000	189.9

Note:

By giving an example of how odds ratios are calculated, the value of 211.5 related to barriers to entry is based on the ratio of the following two odds:
the likelihood of the Authority's intervention with respect to the likelihood of the Authority's authorization, when barriers to entry occur: 1.23:1 (ratio between 55.24% and 44.47% in Table 3);
the likelihood of the Authority's intervention with respect to the likelihood of the Authority's authorization, when barriers to entry do not occur: 0,006:1 (ratio between 0.58% and 99.42% in Table 3).
By comparing the likelihood of the two events (Authority intervention vs. authorization) in two different situations (occurrence vs. non-occurrence of barriers to entry), we find that the likelihood the Authority will challenge a merger when barriers to entry occur is 200 times the likelihood of challenging a merger when barriers to entry do not occur. This outcome, of course, does not allow for the influence of other factors on the Authority's decision.

²⁹ "[T]he odds of an event is the ratio of the expected number of times that an event will occur to the expected number of times it will not occur". Allison (1999), p.9. With respect to probabilities, the use of odds is associated with a more sensitive scale for multiplicative comparisons: they range from 0 to infinitive, while probabilities range from 0 to 1. It means that if one event has a 30% probability to occur and another event a probability of 60%, the latter is twice the probability of the former; but if the likelihood of an event is 60%, no other event could occur at twice that probability.

³⁰ The *odds ratio* is a way of comparing whether the probability of a certain event is the same for two categories. Here we compare an event (the Authority's intervention) in two different situations (e.g., the occurrence or not of barriers to entry). Odds ratios are directly related to parameters in the bivariate *logit* model of Authority's decisions, where the value of its parameters are directly comparable to the outcome of the univariate approach (see the section of this paper on the multivariate approach).

We do not make any further comments on these statistics, herewith introduced since they are directly related to the estimated econometric parameters in the logit model of the Authority's decisions (*infra*). We only observe that the figures in Table 11 are the "*unadjusted*" odds ratios, which do not take account of the joint effect of the above-commented explanatory variables. Indeed, the parameters of the *logit* model are the "*adjusted*" odds ratios, permitting an unbiased estimate of the likelihood of the Authority's enforcement action which allows for the entire set of factors affecting the antitrust decisions.

3.3 Market scenarios

In order to obtain some useful benchmarks for the analysis of decision outcomes, we have to cross-examine those factors that mostly seem to influence the Authority's decisions on merger control activity: barriers to entry, post-merger market shares of the merging parties, increases in the HHI concentration index and possible coordinated effects as a consequence of the merger.

First of all, high post-merger market shares and evidence of barriers to entry are most likely to trigger the Authority's intervention (Table 12). In further detail, with market shares in the range 70-100%, the likelihood of challenging the merger is 95.3% in the presence of barriers to entry, and as low as 3.5% in their absence. The values of these probabilities decrease, respectively, to 65% and 5% in the 40-70% range of market shares, and to 23% and 0.27% for market shares lower than 40%.

TABLE 12
AUTHORITY INTERVENTIONS BY POST-MERGER MARKET SHARES OF MERGING PARTIES AND OCCURRENCE OF BARRIERS TO ENTRY 1995-2003 (% DISTRIBUTION OF NO. OF OBSERVATIONS)

Post-merger market shares		Authority intervention		
		Yes	No	Total
No barriers to entry	0-40%	0.27%	99.73%	100%
	40-70%	5.12%	94.88%	100%
	70-100%	3.50%	96.50%	100%
Barriers to entry	0-40%	23.08%	76.92%	100%
	40-70%	65.71%	34.29%	100%
	70-100%	95.35%	4.65%	100%
Total		1.71%	98.29%	100%

The occurrence of barriers to entry also seems relevant if we examine the distribution of the likelihood of Authority intervention through the change of concentration levels directly induced

by the notified merger (Table 13). The incidence of enforcement action passes from 11% for variations of the HHI index lower than 100 points, to 84% for HHI variations between 500 and 2,000, and up to 100% for HHI changes over 2,000 points. When barriers to entry have not been detected, the likelihood of Authority intervention is lower than 15% for any variation in the HHI index.

TABLE 13
AUTHORITY INTERVENTIONS BY CHANGE IN HHI AND OCCURRENCE OF BARRIERS TO ENTRY 1995-2003 (%
DISTRIBUTION OF NO. OF OBSERVATIONS)

Change in HHI	Authority intervention		
	Yes	No	Total
0	0.52%	99.48%	100%
0-100	0.04%	99.96%	100%
100-500	1.80%	98.20%	100%
No barriers to entry 500-1000	3.85%	96.15%	100%
1000-2000	14.29%	85.71%	100%
≥ 2000	7.14%	92.86%	100%
0	16.67%	83.33%	100%
0-100	11.11%	88.89%	100%
100-500	79.17%	20.83%	100%
Barriers to entry 500-1000	84.21%	15.79%	100%
1000-2000	84.21%	15.79%	100%
≥ 2000	100.00%	0.00%	100%
Total	1.71%	98.29%	100%

The alleged collective dominance hypothesis concerns very few cases in the 1995-2003 period. Authority intervention occurred in four out of a total of five markets where coordinated effects of the concentration were found. Those are markets with quite a high level of concentration (more than 70%), as measured by the CR3 index (Table 14).

TABLE 14
AUTHORITY INTERVENTIONS BY UNILATERAL OR COORDINATED EFFECTS AND CR3 MARKET CONCENTRATION
INDEX 1995-2003 (NO. OF OBSERVATIONS)

CR3	Authority intervention		
	Yes	No	Total
0-40%	68	5,108	5,176
Unilateral effects 40-70%	23	1,102	1,125
70-100%	23	591	614
0-40%	-	-	-
Coordinated effects 40-70%	-	1	1
70-100%	4	-	4
Total	118	6,802	6,920

The following two tables summarize the cross-distribution of observations with respect both to the post-merger market shares of the merging parties and the change in the HHI index, without (Table 15) and with (Table 16) the occurrence of barriers to entry. Moving from one scenario (without barriers, 6,920 observations) to the other (with barriers, 143 observations), we observe a significant change in the clustering of observations, from low levels of both variables to the highest ranks.

TABLE 15
AUTHORITY DECISIONS (INTERVENTION AND NON-INTERVENTION) BY POST-MERGER MARKET SHARES OF MERGING PARTIES AND CHANGE IN HHI 1995-2003 (% DISTRIBUTION OF NO. OF OBSERVATIONS)

Post-merger market shares	Change in HHI					Total
	0-99	100-199	200-499	500-999	≥ 1000	
0-19%	72.24%	2.14%	-	-	-	74.38%
20-39%	10.53%	2.34%	4.41%	0.35%	-	17.63%
40-59%	2.66%	0.30%	0.62%	0.61%	0.29%	4.48%
60-79%	0.78%	0.03%	0.07%	0.04%	0.29%	1.21%
80-100%	1.69%	0.01%	0.03%	0.07%	0.49%	2.30%
Total	87.90%	4.83%	5.13%	1.07%	1.07%	100%

TABLE 16
AUTHORITY DECISIONS (INTERVENTION AND NON-INTERVENTION) BY POST-MERGER MARKET SHARES OF MERGING PARTIES AND CHANGE IN HHI, IN THE PRESENCE OF BARRIERS TO ENTRY 1995-2003 (% DISTRIBUTION OF NO. OF OBSERVATIONS)

Post-merger market shares	Change in HHI					Total
	0-99	100-199	200-499	500-999	≥ 1000	
0-19%	23.08%	-	-	-	-	23.08%
20-39%	8.39%	2.10%	11.19%	0.70%	-	22.38%
40-59%	3.50%	2.10%	3.50%	6.29%	2.10%	17.48%
60-79%	6.29%	-	0.70%	0.70%	3.50%	11.19%
80-100%	2.80%	0.70%	0.70%	3.50%	18.18%	25.87%
Total	44.06%	4.90%	16.08%	11.19%	23.78%	100%

It is clearly evident that the occurrence of barriers to entry is associated with a more concentrated market structure. As a result, the likelihood for a merger to be challenged by the Competition Authority consistently increases (Table 17). More specifically, every case with a

post-merger market share greater than 60% and a change in the HHI concentration index greater than 200 points triggered Authority intervention.

TABLE 17
AUTHORITY INTERVENTIONS BY POST-MERGER MARKET SHARES OF MERGING PARTIES AND CHANGE IN HHI, IN
THE PRESENCE OF BARRIERS TO ENTRY 1995-2003

(% distribution of no. of observations)

Post-merger market shares	Change in HHI					Total
	0-99	100-199	200-499	500-999	≥ 1000	
0-19%	3.03%	-	-	-	-	3.03%
20-39%	-	66.67%	75.00%	-	-	43.75%
40-59%	60.00%	66.67%	100.00%	66.67%	66.67%	72.00%
60-79%	33.33%	-	100.00%	100.00%	100.00%	62.50%
80-100%	75.00%	100.00%	100.00%	100.00%	100.00%	97.30%
Total	15.87%	71.43%	82.61%	75.00%	97.06%	55.24%

(no. of observations)

Post-merger market shares	Change in HHI					Total
	0-99	100-199	200-499	500-999	≥ 1000	
0-19%	1	-	-	-	-	1
20-39%	-	2	12	-	-	14
40-59%	3	2	5	6	2	18
60-79%	3	-	1	1	5	10
80-100%	3	1	1	5	26	36
Total	10	5	19	12	33	79

4. The multivariate approach: the Authority's decision model

Moving from the univariate to the multivariate approach allows us to single out the specific contribution of every factor examined so far with respect to the Authority's final decision.

We have resorted to the bivariate *logit* regression model, with a dichotomous response (dependent) variable. It assumes the value of 1 when the Authority challenges the merger, 0 when the concentration is cleared, even at the conclusion of a formal investigation phase³¹.

The estimates of the regression parameters are shown in Table 18, where a number of diagnostic tests have been reported, mainly based on statistics measuring predictive power³².

³¹ Both the response and the explanatory variables are defined in the preceding sections of this paper.

TABLE 18
BINOMIAL LOGIT MODEL: LIKELIHOOD OF AUTHORITY INTERVENTION ON MERGER CASES 1995-2003

	Model 1	Model 2	Model 3	Model 4	Model 5
Constant term	-6.7764 (*)	-7.2065 (*)	-11.2091 (*)	-18.4094 (*)	-11.0503 (*)
Post-merger market shares of merging parties	0.0404 (*)	0.0352 (*)	1.5357 [§] (*)	1.7667 [§] (*)	1.8450 [§] (*)
Change in HHI	0.000763 (*)	0.000960 (*)	0.3077 [§] (*)	0.2783 [§] (*)	0.2695 [§] (*)
Barriers to entry (dummy)	4.0054 (*)	4.0961 (*)	3.8863 (*)	4.2300 (*)	4.2817 (*)
Buyer power (dummy)	-3.0179 (*)	-1.2270	-1.8047(***)	-2.2653 (**)	-2.3848 (**)
Scope and vertical effects (dummy)	2.0729 (*)	2.2746 (*)	1.9435 (*)	2.6499 (*)	1.6701 (*)
Supra-national geographic market (dummy)		-2.6598 (**)	-2.7432 (*)	-2.3714 (**)	-3.2652 (*)
Sub-national geographic market (dummy)		1.5505 (*)	1.0551 (*)	0.8132 (**)	
Coordinated effects: dummy x CR3	0.0852 (**)	0.0916 (**)	0.0722 (**)	0.0900 (*)	0.0789 (*)
Distribution markets (dummy)		-1.6014 (*)	-1.8342 (*)	-1.6942 (*)	
t (time trend)				2.8079 (*)	
t ² (quadratic time trend)				-0.2681(*)	
No. of Authority interventions $t-1$					-0.2516 (*)
DIAGNOSTIC STATISTICS					
No. of observations	6,920	6,920	6,920	6,920	6,920
R ²	0.0997	0.1053	0.1085	0.1142	0.1084
Maximum-rescaled R ²	0.6288	0.6637	0.6840	0.7200	0.6835
- 2 Log L	467.887	425.246	400.312	355.822	400.981
% of Authority interventions correctly predicted by the model [¶]	57.6%	57.6%	61.9%	67.8%	65.3%

Notes:

§ Logarithmic transformation of the explanatory variable.

(*) The estimated parameter is significantly different from 0 for 1% two-tail test.

(**)The estimated parameter is significantly different from 0 for 5% two-tail test.

(***)The estimated parameter is significantly different from 0 for 10% two-tail test.

¶ The response variable is considered 'Authority intervention' for an estimated probability value greater than 50%. The reported statistics only refer to correctly predicted Authority interventions, not to non-intervention events. With reference to model 5, the overall predictability is 99.2%, detailed as follows:
correctly predicted Authority interventions (prohibitions or authorizations with commitments): 65.3%;
correctly predicted Authority non-interventions (authorizations without commitments): 99.8%.

³² The predictive power is measured by a generalized coefficient of determination (*Max-rescaled R²*, which divides the original *R²* by its upper bound), adjusted to take account of the dichotomous nature of the dependent variable; as such, the unadjusted *R²* can achieve a maximum value less than one. Within the generalized linear models, the *R²* statistic is based on the comparison between the likelihood (*L*) of the intercept-only model and the likelihood of the full specified model. The difference between these two statistics – expressed as $-2\log(L)$ – has a χ^2 distribution. Table 18 shows the statistic $-2\log(L)$ of the specified model, which permits us to assess the goodness-of-fit of various model specifications: lower values of this statistic correspond to more desirable models. The predictive performance of the model has also been assessed in terms of the proportion of correct predictions, both within and outside the sample (*cross-validation*). Table 18 reports the within-sample statistic, with reference to the decision to challenge the merger.

All the explanatory variables illustrated in the previous sections enter the model with the expected sign and are statistically significant (model 1). The concentration level variable plays a significant role only when the coordinated behaviour of the main firms remaining after the concentration has been assumed. In the other cases, the change in the concentration levels as a consequence of the merger seems relevant, proxied by the variation in the HHI index.

The predictive power of the model slightly improves when the geographic sub-national dimension of the relevant markets is considered (model 2), and the distribution market *dummy* variable as well. But we will see later on how both these effects vanish through a more correct specification of the model.

The logarithmic transformation of post-merger market shares and of the change in HHI assumes a non-linear relationship between those variables and the response variable, and further improves the model predictive power (model 3). In other terms, and with respect to the linear specification, the marginal effect of these variables on the likelihood of an enforcement action decreases as the value of these same variables increases³³.

A further improvement is produced by considering a quadratic time trend, able to predict the 1999-2000 peak of Authority interventions followed by quite a sharp decline (model 4). However, even this effect weakens if we properly take account of two dynamic factors associated with the longitudinal characteristic of the used data set.

The first dynamic factor derives from the assumption of independence of the observations – i.e., the outcome for each unit of analysis is unrelated to the outcome for every other observation. This is a very limiting and unrealistic hypothesis when dealing with longitudinal data, and in this field even more so. Indeed, consistency with respect to past decisions on a market is required when the Authority takes another decision on the same market. From an econometric point of view, we can define *clusters* of related observations – in this context corresponding to the definition of the product market – yielding unbiased estimates of the regression parameters (Allison, 1999, Ch. 8)³⁴. The new model estimate (model 5) completely removes any statistical significance from the quadratic time trend variable, as well as from the *dummies* related to the distribution markets and to the sub-national dimension of the relevant markets.

The second dynamic aspects of the Authority's decision model is determined by a "deterrence" (or "dissuasion") effect, proxied by the number of mergers challenged by the Authority in the previous period. In other terms, a high frequency of enforcement actions towards mergers

³³ Federal Trade Commission (2004), p. 15, reports a similar finding: "[t]he resulting properties of the log-transformation are more desirable than a model measuring the relevant variables in levels and also seemed to fit the data much better".

³⁴ As many as 2,442 *clusters* have been defined, out of a total of 6,920 observations, which correspond to 2,442 distinct product markets. The maximum size of a cluster contains 477 observations. We can observe that one of the Federal Trade Commission studies also resorts to such a method of estimation, where clusters correspond to merger cases: "[o]ur procedure uses clustered standard errors, because many mergers involved multiple overlaps. The clustering procedure allowed for the relationship among the related observations". (Coate and Ulrick, 2005, p. 15, footnote 27) . As we argue above, we retain that the use of the product market to define clusters is a more correct choice, since decisions on single product markets have effects that go beyond the single case, becoming a "precedent". On this point see also Voigt and Schmidt (2005), p. 3.

dissuades the notification of likely anti-competitive merger projects. This effect has been proxied by the number of Authority interventions adopted one year previously. It takes the expected sign and an acceptably significant statistical level (model 5)³⁵.

The parameter estimates of a *logit* model can be interpreted in terms of *odds ratios*³⁶. This allows for a direct comparison of the multivariate model outcomes with similar statistics calculated through the univariate approach (Table 19).

The prominent role of barriers to entry is confirmed, even though at a lower level with respect to the univariate analysis. Instead, buyer power plays a stronger role in orientating the Authority's decision than appeared from the analysis of raw data, i.e., not controlled for the effect of the entire set of explanatory variables entering the model. A similar effect presents itself with reference to the supra-national dimension of the relevant market.

TABLE 19
LIKELIHOOD (*ODDS RATIO*) OF AUTHORITY INTERVENTION WITH RESPECT TO NON-INTERVENTION DECISIONS
1995-2003

Characteristics of market structure	Odds ratio	
	Univariate analysis (a)	Multivariate analysis (b)
Barriers to entry (presence of)	211.5	72.4
Buyer power (presence of)	0.52	0.092
Scope and vertical effects (presence of)	10.8	5.3
Geographic market size (w.r.t. the national dimension):		
supra-national	0.10	0.038

Notes:

(a) See Table 11.

(b) Exponential transformation of 'model 5' parameters in Table 18 ($odds\ ratio = \exp^{parameter}$).

5. Simulation of the model and thresholds triggering antitrust intervention

The model estimated in the last section can be used to predict the likelihood of Authority interventions, depending on a number of market scenarios. The parameter estimates are only valid within the sample interval of the data values and, as a consequence, simulation outcomes will refer to feasible market scenarios, i.e., those referred to in Tables 15-17. The results of the model simulations are reported in Table 20, based on parameter estimates of model 5 in Table 18.

³⁵ With reference to other possible temporal effects influencing the Authority's decisions, findings from unreported regressions seem to exclude a change of regime in the law enforcement on merger control, as related to the change of the Authority's Chairman (Giuliano Amato for the 1995-1997 period, and Giuseppe Tesaro for the subsequent analysed years) and Board. A Chairman/Board "dummy" variable does not show any significance when the two dynamic factors mentioned above in the text are properly taken into account.

³⁶ See footnotes 29 and 30 for the definition of "odds" and "odds ratios". Allison (1999), p. 29, refers to "adjusted" *odds ratio*, as they account for the effects of all the other factors entering the *logit* model.

Firstly, an enforcement action by the Authority (with a probability level greater than 50%) seems viable only when barriers to entry occur. This is so independently of any value of the post-merger market shares of the merging parties and of the increment of the concentration levels (Table 20a). The likelihood of Authority's intervention increases significantly for post-merger market shares higher than 40% (Table 20b)³⁷, and it further increases as HHI changes become greater. More specifically, it is more likely that a merger will be challenged (70% probability) when market shares and HHI increments are greater than, respectively, 40% and 500 points, or with a 60% post-merger market share for any change in the HHI index.

The probability of challenge is even higher when barriers to entry and vertical effects of the concentration occur jointly. In such a case, probabilities greater than 70% are associated with quite low post-merger market shares (greater than 20%, see Table 20c).

Finally, buyer power plays an important role. It is always able to countervail the occurrence of barriers to entry, except when the existence of vertical or horizontal effects of the concentration strengthens the anti-competitive effects deriving from entry barriers. In the latter case, the Authority's intervention is very likely for post-merger market shares greater than 60% and changes in the HHI index over 200 points (Table 20e).

³⁷ By comparing the data of Tables 20b and 17, we may note that what is represented in the upper right of Table 20b (data within the area delimited by the dotted line) does not reflect a situation observed in reality (empty cells in Table 17).

TABLE 20
LIKELIHOOD OF AUTHORITY INTERVENTION THROUGH THE SIMULATION OF ALTERNATIVE MARKET SCENARIOS (1995-2003)

TABLE 20A

BASE SCENARIO: NO BARRIERS TO ENTRY, BUYER POWER OR SCOPE/VERTICAL EFFECTS

		Change in HHI				
		0-99	100-199	200-499	500-999	≥ 1000
Post-merger market shares of merging parties	0-19%	0.11	0.15	0.19	0.24	0.26
	20-39%	0.70	0.99	1.24	1.52	1.65
	40-59%	1.77	2.49	3.10	3.79	4.10
	60-79%	3.24	4.53	5.62	6.83	7.36
	80-100%	5.11	7.09	8.73	10.54	11.33

TABLE 20B
BARRIERS TO ENTRY

		Change in HHI				
		0-99	100-199	200-499	500-999	≥ 1000
Post-merger market shares of merging parties	0-19%	6.95	9.42	11.36	13.42	14.30
	20-39%	32.55	40.73	46.08	51.12	53.12
	40-59%	55.22	64.17	69.11	73.35	74.94
	60-79%	69.62	77.12	80.82	83.86	84.96
	80-100%	78.60	84.49	87.20	89.36	90.14

TABLE 20C
BARRIERS TO ENTRY AND SCOPE/VERTICAL EFFECTS

		Change in HHI				
		0-99	100-199	200-499	500-999	≥ 1000
Post-merger market shares of merging parties	0-19%	24.82	30.75	34.64	38.34	39.82
	20-39%	70.38	77.60	81.19	84.13	85.20
	40-59%	86.29	90.37	92.15	93.54	94.03
	60-79%	92.22	94.69	95.71	96.49	96.77
	80-100%	95.04	96.65	97.31	97.81	97.98

TABLE 20D
BARRIERS TO ENTRY AND BUYER POWER

		Change in HHI				
		0-99	100-199	200-499	500-999	≥ 1000
Post-merger market shares of merging parties	0-19%	0.72	1.02	1.27	1.55	1.68
	20-39%	4.47	6.21	7.66	9.25	9.95
	40-59%	10.64	14.46	17.46	20.65	22.02
	60-79%	18.09	23.94	28.27	32.66	34.49
	80-100%	26.13	33.61	38.79	43.83	45.87

TABLE 20E
BARRIERS TO ENTRY, BUYER POWER AND SCOPE/VERTICAL EFFECTS

		Change in HHI				
		0-99	100-199	200-499	500-999	≥ 1000
Post-merger market shares of merging parties	0-19%	3.63	5.02	6.16	7.41	7.95
	20-39%	19.49	25.55	29.95	34.37	36.20
	40-59%	38.09	46.92	52.48	57.59	59.58
	60-79%	53.25	62.34	67.43	71.83	73.48
	80-100%	64.56	72.76	76.96	80.46	81.74

Notes:

Simulations based on 'model 5' parameters in Table 18. The value of the variable 'No. of Authority interventions t_{-1} ' has been put to its average value within the estimation period (4.66). Probability values reported in this table are simple averages of the probabilities estimated at each single value of the explanatory variables. Finally, data in the upper right of the area delimited by the dotted line do not reflect a situation observed in reality (see Tables 15 to 17).

6. Conclusions

The findings presented in this paper seem relevant both in terms of content as well as of methodology, suggesting a few paths to follow in future research.

On the first point, the consistency analysis of Authority decisions carried out in this paper adds to the transparency and predictability of the merger control enforcement action in Italy. The relevance of traditional market indicators has been highlighted, and the significance of the multivariate approach confirmed. According to this, decisions are determined by multivariate market scenarios – as represented by the occurrence of entry barriers, vertical and scope effects, buyer power, ease of collusion, etc. – which contribute to setting thresholds in terms of market shares and market concentration index levels and variations, and to predicting the likelihood of Authority intervention thereof.

Barriers to entry play a fundamental role in depicting the various scenarios, and in determining decisions as well. The methodological approach followed in this paper has focused on the presence of the concept within the Authority's resolution. However, the importance of this variable should suggest an in-depth study of its nature³⁸, even to the point of verifying that their use is consistent within the examined cases³⁹.

The treatment of the dynamic aspects of the decision model – in this paper, as well as in other papers on the same subject matter – is not yet completely satisfactory. The efficiency of the parameter estimates of the model definitely improves by identifying clusters of observations, which take into account the non-independence of the observations within each cluster. But this is not the entire story. The weakness of the independence hypothesis is much stronger when referring to case law decisions taken by the same Authority, which tend to assure their consistency with the previous ones. In this view, a more accurate specification of the time frame of the model, as well as an explicit consideration of some sort of "precedent effect", would yield more robust results.

Finally, a methodological comment. The use of lexical-textual methods of analysis in this field is quite innovative. The findings up to this point seem encouraging,

³⁸ As an example, Voigt and Schmidt (2005), pp. 87ff, distinguish three groups of barriers to entry, according to their origin: (1) state-mandated barriers (such as tariff and non-tariff trade barriers, patents, copyrights, etc.), (2) structural barriers (such as economies of scale, learning by doing, capital requirements, distribution nets, etc.), (3) strategic barriers (such as product differentiation, advertising, goodwill, vertical restraints).

³⁹ The reader may refer, among all, to a few articles appeared in the May 2004 issue of *Papers and Proceedings* of the American Economic Review. One of these papers contends that "economists have not yet been able to reach broad consensus over the definition of an entry barrier, and this probably hindered the development of efficient antitrust policy". McAfee et al. (2004), p. 465.

even though instruments have been used here more as an information retrieval/extraction tool to classify the content of the decisions than as a text-mining and content-analytic instrument proper⁴⁰. The exploratory and analytical effectiveness of the tool utilized allows for further testing in the latter direction.

⁴⁰ "Content analysis is a research technique for making replicable and valid inferences from texts (or other meaningful matter) to the contexts of their use". Krippendorff (2004), p. 18. Distinguishing qualities of content-analytic methods are quantification ("procedures that permit categorized data to be translated into nominal, ordinal or interval scales"), systematicity and procedural transparency ("use of explicitly formulated rules and procedures"), theoretical relevance ("usually not a descriptive method; it seeks theoretical relevance through hypothesis and theory testing"). Lee and Fielding (2004), p. 534.

References

Empirical literature by jurisdiction:

Canada

Khemani, R.S. and Shapiro, D.M. (1993), "An empirical analysis of Canadian merger policy", *The Journal of Industrial Economics*, 41(2), pp. 161-77.

Khemani, R.S. and Shapiro, D.M. (1994), "The administration of Canadian merger policy, 1986-1989", *The Antitrust Bulletin*, 39(3), pp. 771-95.

European Union

Aktas, N., de Bodd, E. and Roll, R. (2004), "European M&A regulation is protectionist", EFA 2004 Maastricht Meetings Paper No. 2318.

Bergman, M.A., Jakobsson, M. and Razo, C. (2005), "An econometric analysis of the European Commission's merger decisions", *International Journal of Industrial Organization*, 23(9-10), pp. 717-37.

Duso, T., Neven, D.J. and Röller, L.H. (2003), "The political economy of European merger control: evidence using stock market data", Discussion Paper FS IV 02-34, Wissenschaftszentrum Berlin.

Lindsay, A., Lecchi, E. and Williams, G. (2003), "Econometrics study into European Commission merger decisions since 2000", *European Competition Law Review*, 24(12), pp. 673-82.

Neven, D., Nuttall, R. and Seabright, P. (1998), "Enforcement of the European merger regulation. The merger decisions of the European Commission", in Philips L. (ed.), *Applied industrial economics*, Cambridge, England: Cambridge University Press.

Neven, D.J. and Röller, L.H. (2002), "Discrepancies between market and regulators: an analysis of the first ten years of EU merger control", in Swedish Competition Authority (ed.), *The pros and cons of merger control*, Stockholm: Swedish Competition Authority.

Verouden, V. (2004), "The role of market shares and market concentration indices in the European Commission's Guidelines on the assessment of horizontal mergers under the EC Merger Regulation", mimeo, FTC/DOJ Workshop on Merger Enforcement, Washington DC.

Voigt, S. and Schmidt, A. (2005), *Making European merger policy more predictable*, Berlin: Springer-Verlag.

Germany

Lauk, M. (2002), "Econometric analysis of the decisions of the German Cartel Office", mimeo, Technische Universität Darmstadt.

New Zealand

Strong, N., Bollard, A. and Pickford, M. (2000), "Defining market dominance: a study of antitrust decisions on business acquisitions in New Zealand", *Review of Industrial Organization*, 17(2), pp. 209-27.

United Kingdom

- Davies, S.W., Driffield, N.L. and Clarke, R. (1999), "Monopoly in the UK: What determines whether the MMC finds against the investigated firms?", *The Journal of Industrial Economics*, 47(3), pp. 263-83.
- Office of Fair Trading (2005), "Ex post evaluation of mergers", A report prepared for the Office of Fair Trading, the Department of Trade and Industry and the Competition Commission by PricewaterhouseCoopers LLP.
- Weir, C., (1992), "Monopolies and Merger Commission, merger reports and the public interest: a probit analysis", *Applied Economics*, 24, pp. 27-32.

United States of America

- Coate, M.B. and McChesney, F.S. (1998), "Enforcement of the US merger guidelines. Empirical evidence on FTC enforcement of the merger guidelines", in L. Philips (ed.), *Applied industrial economics*, Cambridge, England: Cambridge University Press.
- Coate, M.B., Higgins, R.S. and McChesney, F.S. (1995), "Bureaucracy and politics in FTC merger challenges", in Chesney, F.S. and Shughart II W.F. (eds), *The causes and consequences of Antitrust. The public-choice perspective*, Chicago, IL: The University of Chicago Press.
- Coate, M.B. (1996), "Merger enforcement at the Reagan/Bush FTC", in Coate, M.B. and Kleit A.N. (eds.), *The economics of the antitrust process*, Boston, MA: Kluwer Academic Publishers.
- Coate, M.B. and Ulrick, S.W. (2005), "Transparency at the Federal Trade Commission: The horizontal merger review process 1996-2003", Federal Trade Commission, Bureau of Economics, Economic Issues Paper.
- Coate, M.B. (2005a), "Twenty years of Federal Trade Commission merger enforcement activity (1985-2004)", Potomac Paper in Law and Economics 05-02.
- Coate, M.B. (2005b), "An empirical analysis of merger enforcement under the 1992 Merger Guidelines", Potomac Paper in Law and Economics 05-03.
- Coate, M.B. (2005c), "Economic models in merger analysis: a case study of the Merger Guidelines", Potomac Paper in Law and Economics 05-04.
- Federal Trade Commission (2004), "Horizontal merger investigation data, fiscal years 1996-2003", FTC Staff Report, February 2 (revised August 31).
- Kwoka, J. (2004), "Some thoughts on concentration, market shares, and merger enforcement policy", mimeo, FTC/DOJ Workshop on merger enforcement, Washington DC.

Analysis of textual data; Text-Mining:

- Baiocchi, F., Bolasco, S., Canzonetti A. and Capo, F.M. (2005), "Estrazione di informazione da testi per la classificazione automatica di una base documentale: la soluzione di Text Mining per l'Authority della Concorrenza", in Bolasco, S., Canzonetti, A. and Capo F.M. (eds), *Text Mining. Uno strumento strategico per imprese e istituzioni*, Roma: CISU.
- Bolasco, S. (2004), "L'analisi statistica dei dati testuali: intrecci problematici e prospettive", in Aureli Cutillo, E. and Bolasco, S. (eds), *Applicazioni di analisi*

statistica dei dati testuali, Roma: Casa Editrice Università degli Studi di Roma "La Sapienza".

Bolasco, S. (1999), *L'analisi multidimensionale dei dati*, Roma: Carocci,.

Bolasco, S., Baiocchi, F., and Morrone, A. (2004a), *TALTAC Guida all'uso*, Roma: CISU, updates of release 1.6.2 (June), www.taltac.it.

Bolasco, S., Bisceglia, B. and Baiocchi, F. (2004b), "Estrazione automatica di informazione dai testi", *Mondo Digitale*, 3(1), pp. 27-43.

Bolasco, S., Canzonetti, A., Capo, F.M., Della Ratta-Rinaldi, F. and Singh, B.K. (2004c), "Understanding Text Mining: a pragmatic approach", Knowledge Mining – NEMIS 2004 Final Conference, Athens.

Krippendorff, K. (2004), *Content analysis: an introduction to its methodology*, 2nd edition, London: Sage Publications Ltd.

Lee, R.M. and Fielding, N.G. (2004), "Tools of qualitative data analysis", in Hardy, M. and Bryman, A. (eds.), *Handbook of data analysis*, London: Sage Publications Ltd.

Salton, G. (1989), *Automatic text processing: The transformation, Analysis and Retrieval of Information by computer*, Reading, MA: Addison-Wesley Publishing Co.

Logit models:

Allison, P.D. (1999), *Logistic regression using the SAS system. Theory and application*, Cary, NC: SAS Institute Inc.

Maddala, G.S. (2001), *Introductions to econometrics*, 3rd Edition, Chichester, England: John Wiley & Sons Ltd.

Barriers to entry:

Carlton, D.W. (2004), "Why barriers to entry are barriers to understanding", *American Economic Review*, 94(2), pp. 466-70.

McAfee, R.P., Mialon, H.M. and Williams, M.A. (2004), "What is a barrier to entry?", *American Economic Review*, 94(2), pp. 461-65.

Schmalensee, R. (2004), "Sunk costs and antitrust barriers to entry", *American Economic Review*, 94(2), pp. 471-77.

Appendix I - The Authority's database on merger control and the experimental use of an automatic procedure to classify qualitative information

I.1 The Authority's database on merger control

According to Italian law no. 287/90 (the Competition and Fair Trading Act), all concentrations between undertakings whose turnover exceeds specified thresholds require advance notification. If the merger raises concerns over its competition effects on the affected markets, the Authority, within 30 days of the notification, commences an investigation. Within 45 days of the commencement of the investigation, if the Authority deems that the merger creates or strengthens a dominant position to the extent that it substantially reduces competition on a lasting basis, it prohibits it, or authorizes it, provided that the original project is amended in order to remove the distortive aspects. Within 20 days the final decision is published in the Authority's Bulletin.

All data on the projected merger is communicated by a form and, once the concentration has been examined by the Authority and a final decision has been taken, the information of interest on the notified transaction, the affected markets and the firms involved (both the merging parties and their main competitors) is put manually into an electronic database (named GATC, an acronym for "Gestione delle Attività di Tutela della Concorrenza" – Competition Activity Management) by the Authority's employees.

The case is given a referencing code, which identifies the administrative proceeding. The starting date (usually corresponding to the date of notification), the closure date (usually that of the final decision), the date of the possible commencement of an investigation are also registered, as well as the final outcome of the Authority's decision. The latter can be an authorization (no commencement of the investigation phase), or, after a thorough investigation, a prohibition or an authorization subject to commitments taken by (or imposed on) the merging parties. It is important to mention that the operations which have been withdrawn, even after the commencement of the investigation phase, are not registered in the database, since economic information on markets could not have been exhaustively

checked by the Authority's investigative units and, as such, may lack the necessary reliability.

With reference to the analysis carried out in this paper, the response variable of the model (the Authority's intervention) is strictly related to the final outcome (authorization vs. prohibition or authorization with commitments), irregardless of the fact that an investigation phase had started. That means that the response variable is only related to the possible anti-competitive effects on the relevant markets, and not to administrative procedural issues.

The usual "conceptual" definition of the antitrust market herewith applies: the smallest product and geographic area where a hypothetical permanent increase in relative prices would be profitable. Of course, a single merger could affect more than one relevant market, and only in some of them could it raise concerns about anti-competitive effects.

The "operational" definition of the relevant market in the database includes, within each notified transaction:

- the product market: coded according to the ISTAT (the Italian National Institute of Statistics) classification system of economic activities, with a possible extension in order to identify it univocally;
- the geographic market: it takes account of the geographic extension of the product market (national, sub-national, supra-national), and of its specific area code (mainly at the provincial level of coding).

If necessary, even the distribution channel is considered.

The market shares of the merging parties and its four main competitors within each relevant market are registered, usually expressed in terms of the firm's turnover. Market share data allows the calculation of pre- and post-merger market shares of the merging parties, as well as indexes of the market concentration level (mainly concentration ratios). Further data is registered in the database, such as the type of concentration (merger, acquisition, joint venture), as well as demographic information on the companies involved, turnover and their role within the transaction project (target, bidder, participant in the joint venture, etc.).

The following table shows a quantitative view of the database content, from the entire period 1990-2003: 4,566 examined transactions corresponding to 11,738 relevant markets. Sixty cases were formally investigated and 45 were prohibited or authorized subject to commitments, corresponding, respectively, to 599 and 425 markets. Henceforth, on average, 2.57 markets were identified and examined for each case, and as many as 10 were defined when an investigation was performed.

TABLE A.1
MERGERS AND ACQUISITIONS DATABASE 1990-2003

Number of relevant markets in merger cases	Number of merger cases	Number of relevant markets	Number of formally investigated cases	Number of relevant markets in formally investigated cases	Number of formally investigated cases not authorized	Number of relevant markets in formally investigated merger cases not authorized
0	79	0	0	0	0	0
1	2,633	2,633	16	16	10	10
2-4	1,381	3,560	20	54	17	45
5-15	395	3,058	14	115	9	79
16-50	66	1,598	8	187	8	187
50+	12	889	2	227	1	104
Total	4,566	11,738	60	599	45	425

Source: Italian Competition Authority.

I.2 Automatic coding of Authority resolutions

Data entry is a very time-consuming and costly activity, depending both on the complexity of each case, and on the number of cases. The highest level of complexity concerns the investigated mergers, where a great deal of information of interest can be found (e.g., a large number of relevant markets, detailed data on market shares of the merging parties and their competitors, etc.), and quite often only partially contained in the text of the final decision (the original case files have to be read). For all these reasons, a manual coding is necessary.

However, most of the notified mergers do not raise competition concerns (see statistics in the table below), and the content of the final decision can be quite easily coded and entered into the database. Usually, these texts show a number of regularities (in terms of textual structure, use of legal terms and substance), and are well suited for the application of automatic procedure of *information extraction*.

In 2004, the Authority has launched a pilot project (named TAPA which stands for *Trattamento Automatico dei Provvedimenti dell'Autorità* – Automatic Treatment of the Authority's Decisions) to implement a prototypical software application for the automatic coding of decisions related to not-investigated merger cases. The research team developed a customized version of the software programme TALTAC, and it was applied to an homogenous *corpus* of decisions (3,549 documents for the period 1995-2003, not related to the investigated mergers and excluding decisions taken before 1995, due to their not-homogeneous textual structure).

The automatic procedure, build as a set of sequential operations, detects and extracts the following information:

- the various sections of the text of the decision;
- the name and the role (in the merger transaction) of the merging parties, also by considering that the same company can be mentioned in a number of different ways within the same document. All citations (acronym, full name with or without legal form) are univocally identified;
- parent companies;
- every other cited company;
- the product market;
- the geographic market;
- the turnover thresholds (triggering the obligation to notify);
- market shares.

This procedure resorts to: i) external dictionaries, ii) algorithms based on the recognition of word sequences, iii) a set of queries based on wild cards or Boolean operators.

- i) External dictionaries allow for the identification and lexicalisation of the companies' names and of the product market definitions, transforming the sequence of terms into a unique entry.
- ii) The algorithm, which operates according to the rule *<incipit + company's name + legal form>* (Figure 1), identifies and lexicalises other entities *<company>* not detected through the dictionary.

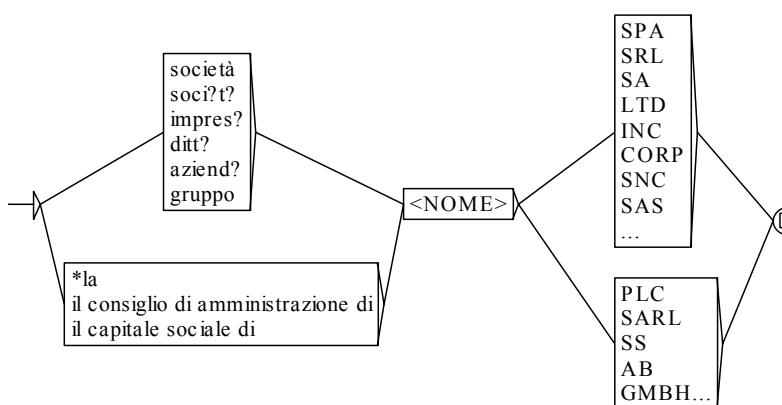


FIGURE 1: Diagram of a rule which detects the name of the company

- iii) Finally, the queries check for the occurrence in the documents of a set of "sensitive" terms, by populating the database fields:

- a. with a certain label (geographic market): e.g., where the sequences "dimension? geografic* LAG15 local*" or "mercat? geografic* LAG15 local*" occur, the database field "geographic market" is coded as "LOCALE" (local);
- b. with lexicalised sequences (corporate groups); e.g., a company name which follows expressions such as "è LAG7 controllat? LAG7 da*" o "partecipat? LAG7 da*" will identify a parent company. By replicating the same procedure, the ultimate parent company can be identified;
- c. with entire sequences of text (product market): e.g., when a product market has not been detected by the dictionary list, the sequence "part? LAG8 attiv? n* settor? d*" OR "part? LAG8 attiv? n* mercat? d*" will detect the incipit (opening words) of a sequence of words which define the market.

Appendix II - Methodologies and tools for textual analysis of the Authority's decisions

II.1 The tool: TALTAC

TALTAC⁴¹ is a software library specifically designed to perform a comprehensive content analysis, from keyword extraction from text, up to information retrieval and information extraction functionalities. It is quite a flexible tool, allowing for a wide range of statistical procedures, interaction with other software programs (such as SPSS, Spad, Lexico ecc.), as well as the development of personalized software applications, giving users ultimate control on their research requirements.

The main functionalities which TALTAC provides are: A) text acquisition, B) text pre-processing, C) lexical analysis, D) information retrieval and extraction, E) textual analysis.

A) Text acquisition and corpus management. As in the case of the Italian Antitrust Authority's decisions, the *corpus* analysed is usually a collection of a (large) number of documents, saved as separately distinct files. TALTAC can access these documents in order to compose a unique *corpus*. Had some categorical information been assigned to each document, it can then be automatically tracked and saved by the software, e.g., in a spreadsheet table. If necessary, the *corpus* can be

⁴¹ The Italian acronym for Automatic Lexico-Textual Treatment for Content Analysis.

partitioned into sub-texts, in order to analyse sub-occurrences with respect to each partition.

B) *Text pre-processing*⁴². Through normalisation and parsing algorithms it is possible, respectively, to clear text of duplicates (standardising the typing of numbers and words, punctuation, symbols, etc.), and convert it in a TALTAC-readable form. Entity recognition, with the use of built-in or user-defined term lists, can also be applied to texts in order to categorize ambiguous words or word sequences – such as surnames, toponyms, grammatical expressions (locutions, adverbs), acronyms, etc. – and to reduce ambiguities in the original text.

C) *Lexical analysis*. It provides for the study of the corpus vocabulary, i.e., of the language. It is a “vertical” analysis, where the representation of its content does not take account of the discourse development; words are extracted like they would be from a ballot-box, which is named “bag of words” in this context. To reconstruct the lexicon from a corpus means the production of statistics on verbs, adverbs, nouns, adjectives.

More specifically, through the use of statistical tools of analysis, lexical analysis allows for the description of some regularities of the language. These are measured by the relative incidence of some classes of words (*imprinting*) which are able to differentiate the original texts, to assess its level and type (the percentage of words from a basic vocabulary, the abstractness/concreteness of the discourse, a positive/negative tone).

D) *Information retrieval and information extraction from text*. “Mining” is an important phase in text analysis, focusing on the most significant part of the vocabulary. The *peculiar language* is identified, i.e., the 12-15% of the vocabulary most important to perform a textual analysis. Two situations can be devised, with or without the use of queries.

The extraction of the peculiar language *without* specific queries can be made with the use of *exogenous* resources. For instance, through the calculation of the standard deviation of the use of words with respect to the frequency of use in a lexicon taken as a benchmark (i.e., in terms of expected values)⁴³. But also resorting to *endogenous* resources (e.g., through the calculation of specificities) we are able to select the particular language in each partition of the text (year, month, issue of the published bulletin, type of decision, etc.).

⁴² Text is “[...] a sequence of symbols, part of a codex (ANSI) which identifies two sets of characters: the delimiters and the non-delimiters, also known as alphabet. The single units of analysis, indeed, are all the sequences of elements of an alphabet among two delimiters”. (Bolasco et al., 2004a).

⁴³ “Extraction of keywords of the corpus by calculation of over/underuse of a term against a frequency lexicon assumed as reference language”; Bolasco et al. (2004a).

By using a query, instead, the calculation of an index such as *TFIDF*⁴⁴ (Salton 1989) allows for the selection of terms closer to the request, producing a ranking of retrieved documents by relevance.

Once the *corpus* has been loaded, the text pre-processing is done and the vocabulary with the occurrences is calculated, the user can perform information retrieval and information extraction from text, independently of any research hypothesis (as described below), through concordance analysis⁴⁵ or complex sets of queries which support Boolean operators. If documents are partitioned into sections (as in the case-study of the Authority's decisions), queries can be applied to any one of them selectively. Queries are also useful for extracting a subset of documents in which a searched term (or a string of words) is present, by creating a new textual variable which tags the presence/absence of the term within the documents. This same textual variable can assume different values, and is automatically saved in a table created into TALTAC. This table, as well as many others created during any user's working session, is saved in a "session database" in TALTAC. By accessing the database, the user can export tables, lists of words (as the result of the analysis) or the entire *corpus* (lexicalised by personalized lists or semantically categorised).

E) *Textual analysis* is related more directly to the analysis of the corpus, specifically through more or less complicated concordance analyses, globally through co-occurrence analyses. The latter are derived: *directly*, from the statistical analysis of sequences (leaders/followers by a predetermined LAG operator; see below examples concerning the absence of barriers to entry) with respect to pivot words; *indirectly*, by the identification of latent semantic dimensions provided through techniques of data reduction: correspondence analysis, singular value decomposition, multidimensional scaling.

When use is not made of multidimensional statistical methods, textual analysis allows us: i) to solve complex queries (analysis of concepts) by extracting from the corpus the most relevant documents through the *TFIDF* index; ii) to visualize the *named entities* searched; iii) to categorize text fragments by creating new "textual" variables, which populate the field of a structured database, just as in the case study analysed below.

⁴⁴ The *TFIDF* index (Salton, 1989) is $w = tf \cdot \log N/n$, where *tf* is the frequency of a term within every document, *n* represents the number of documents which contain that term, and *N* is the entire number of documents in the corpus. Terms are weighted according to their relevance, and the index is higher as terms are more frequent in a small number of documents.

⁴⁵ Concordance analysis is defined as the "systematic study of local contexts of a selected word" in a text, where "local left/right co-texts" are "[...] a set of words closest to a fixed term which works as a pole (or pivot)". See Bolasco (1999), p. 184. Many of the concepts recalled in this appendix are borrowed from the cited reference.

II.2 The research strategy: from text to data

Using the TALTAC text mining software program, qualitative data has been extracted from the text not currently codified in the Authority's document warehouse and in the other statistical and economic databases managed by the Authority.

The following is an overview of the strategy pursued in order to build a matrix of qualitative data values from the text. The matrix has been used for statistical analysis and econometric estimation of the Authority's decision model related to the control of mergers and acquisitions between companies.

The unit of analysis is the set of "lexical profiles" which arise from the internal structure of the *corpus* analysed⁴⁶. The conventional structure (sections) of every single document to which the procedures of information extraction are applied is the following:

- *Preambolo* (Preamble)
- *Le parti* (The merging parties)
- *Descrizione dell'operazione* (Operation description)
- *Qualificazione dell'operazione* (Operation qualification)
- *Valutazione della concentrazione* (Evaluation of the concentration)
- *Delibera (dispositivo)* (Resolution)

Most of the relevant information about the concentration is actually contained in the fifth section (*Evaluation of the concentration*), which represents the *contextual unit* for the purposes of the automatic analysis of the *corpus*⁴⁷.

Attention has focused on the following topics:

- entry barriers
- buyer power
- vertical effects
- coordinated effects

and the text-mining research strategy follows these steps:

normalisation

- ❑ concordance analysis
- ❑ fusions
- ❑ queries
- ❑ checking of the results

⁴⁶ *Id.*, p. 183.

⁴⁷ *Id.*, p. 208ff.

A brief description of each step is given. For further details see the references cited in the bibliography.

Applying certain algorithms to the *corpus*, the **normalisation step** aims at reducing the heterogeneity of the text, standardising the way in which words are written.

Then, starting from a pivotal term strictly related to a concept of interest (e.g., the word “barriers” to find any references as to the presence/absence of market entry barriers within the documents), the **concordance analysis** provides a comprehensive listing of a given item in a *corpus* also showing its immediate context, in order to obtain the specific contexts of use in which the pivotal term has been cited (Figure 2).

The screenshot shows the 'Concordanze' window with the search term 'barriere' entered. The results are displayed in a table with columns for 'Forma grafica', 'Occ', 'IDFra...', 'Intorno sinistro', 'Forma grafica', and 'Intorno destro'. The table lists various occurrences of the word 'barriere' in different contexts, such as 'interessata sia della circostanza che non esistono notevoli', 'Tenuto conto della circostanza che non esistono notevoli', and 'Italia sia della circostanza che non esistono notevoli'.

Forma grafica	Occ	IDFra...	Intorno sinistro	Forma grafica	Intorno destro
barriera	36	P9482	interessata sia della circostanza che non esistono notevoli	barriere	all' entrata in tale settore , e considerata la presenza
barriere	607	P9537	Tenuto conto della circostanza che non esistono notevoli	barriere	all' entrata in tali mercati , l' operazione in esame
Barriere	1	P9830	Italia sia della circostanza che non esistono notevoli	barriere	all' entrata in tale settore , e considerata la presenza
Barriera	4	P10145	dalla ED'S Italia , nonché dell' assenza di notevoli	barriere	all' entrata in tale settore , caratterizzato dalla
		P12508	quota marginale , nonché dell' assenza di notevoli	barriere	all' entrata in tale settore caratterizzato dalla presenza
		P4238	del cioccolato da copertura non presenta particolari	barriere	all' entrata in quanto sia le materie prime sia la
		P5282	consideri , inoltre , che non sono presenti particolari	barriere	all' ingresso nel mercato del City Courier , ove l'
		P5283	operazione di concentrazione non presenta particolari	barriere	all' ingresso . Fattori quali la dimensione societaria
		P5460	sviluppo maturo , dove non sono presenti particolari	barriere	all' ingresso . A riprova di ciò , le parti segnalano
		P5555	mercati , infine , non sembrano presentare particolari	barriere	all' ingresso né di natura economica , né di natura
		P4717	caratteristiche del settore che non presenta particolari	barriere	all' entrata e soprattutto del fatto che , stante la
		P6035	ammontanti all' 84% del mercato , non vi sono particolari	barriere	all' entrata . In particolare , non vi è la necessità
		P6355	consideri , inoltre , che non sono presenti particolari	barriere	all' ingresso nel mercato del recapito urbano urgente
		P6414	In ogni caso si rileva che non vi sono particolari	barriere	all' entrata in tale mercato , sia con riferimento
		P6414	in tali mercati non sembrano sussistere particolari	barriere	all' entrata .
		P6612	mercato interessato non presenta , infatti , particolari	barriere	all' entrata né di carattere economico , né di tipo

FIGURE 2: Example of concordance analysis for the pivot term “barriere” (barriers)

From the list of concordances (Figure 2) a frequency table is created, **fusing** (merging) the words adjacent to the pivotal term and counting the number of occurrences. The analysis of the resulting frequency table permits us, in turn, to find the most frequent *incipit* (i.e., opening words), giving evidence of those expressions actually related to the topic of interest and, among them, the most frequent ones. It also indicates which words were actually used in the analysed text.

Such information represents the basis for the formulation of the **queries** to be used for document selection and retrieving. The queries in TALTAC are formulated in terms of regular expressions following a syntax which supports Boolean operators. At the same time, the different “values” of the qualitative variables – the presence or absence of the terms within the documents – are created and saved in the table of fragments.

The final step of the research strategy is aimed at **verifying** the ability of the queries to retrieve the topic of interest with reasonable coverage. In the field of information retrieval, the effectiveness of the search tool is evaluated by resorting to *precision* and *recall* indexes⁴⁸.

These two indexes are usually negatively correlated; therefore, the queries can be enriched (using synonyms, domain terminology, etc.) to obtain a higher recall. Nevertheless, the “noise” will unavoidably increase going against precision, which will lead to an increase in the percentage of non-relevant documents retrieved. On the contrary, the strategy pursued in this work tends to minimise the “noise”, increasing the *precision* and reducing the *recall*. Such a strategy avoids the risk of classifying non-pertinent concepts with regard to those which are relevant for the antitrust assessment of mergers.

To give an example of how documents are retrieved, let us consider one query (in Italian) actually used for the definition of the variable *barriers*:

"non LAG4 presen LAG3 barriere" OR "inesisten* LAG2 barriere"*

The LAG n operator, put between two words, implies that the query refers to all the expressions containing the two words separated by other n words. With the use of the character *star* (*) the query retrieves words that end in different ways (for example, “presen*” retrieves words like “presenti” and “presenza”).

The query written above is formed by two *subqueries* (or *nested queries*), linked to each other by the logical operator “OR”, which allows for alternative expressions within the same unique query. As a matter of fact, the first *subquery* finds expressions such as:

- “non risultano essere presenti particolari barriere”
- “non sono presenti rilevanti barriere”
- “non presenta barriere”.

Instead, the second *subquery*:

- “inesistenza di particolari barriere”
- “inesistenza di rilevanti barriere”
- “inesistenza di barriere”.

Other Boolean operators and wild cards (OR, AND, AND NOT, *, ?, delimiters) can be used to build more complex queries.

⁴⁸ According to logical reasoning, results emerging from the application of queries to a set of documents for information retrieval purposes are the following:

- a subset A of relevant and retrieved documents (also said “true positive” – TP);
- a subset B of relevant but not retrieved documents (“false negative” – FN);
- a subset C of not-relevant but retrieved documents (“false positive” – FP);
- a subset D of not-relevant and not-retrieved documents (“true negative” – TN).

The *precision* index is the fraction of retrieved documents that are relevant (in symbols $A / (A+C)$), while the *recall* index is the fraction of relevant documents that are retrieved (in symbols $A / (A+B)$).

TALTAC uses a specific module, namely “*Ricerca di Entità*” (RE, the Italian for Regular Expressions), for query execution and for assigning values to the textual variables (Figure 3).

FIGURE 3: TALTAC’s Analysis of the Named Entities module

The formulation of a query in the Analysis of the RE module needs to be given some specific information: which section of the document has to be searched (“*VALUTAZIONE DELLA CONCENTRAZIONE*” – evaluation of the concentration – as in the example), the name of the textual variable that will be created (“*BARRIERE*” – barriers), and the value automatically assigned to the variable in the database associated with that query (“NO”).

Once the search has been performed, the document section searched and the documents recalled are highlighted. In the following example (Figure 4), the query detected the text string “*non presenta particolari barriere*” in the document section “*VALUTAZIONE DELLA CONCENTRAZIONE*”.

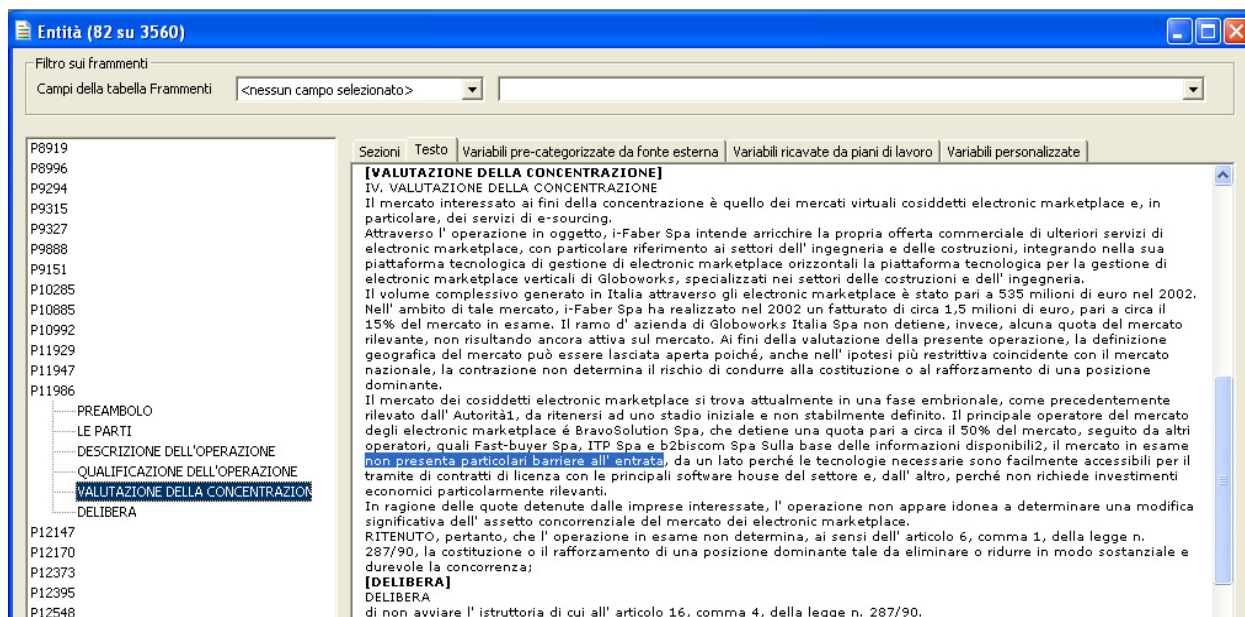


FIGURE 4: Text string identified by a query which has been applied to the "Evaluation of the concentration" section of the Authority's documents

By applying the same search procedure (with different queries) to each topic of interest, a **matrix of textual data** (see below) is yielded, which associates an entire set of variables – registering the presence/absence of the topic under consideration in the query ("SI", yes, "NO", with a blank cell meaning "no occurrence") – with every examined merger operation. In this example, different values of the topics (concepts) of interest "entry barriers", "buyer power", and "vertical effects" are assigned to each antitrust case.

CASE REF.	ENTRY BARRIERS	BUYER_POWER	VERTICAL EFFECTS
C9836	NO		
C9842		SI	
C9849			SI
C9850	SI		
C9852	SI		
C9853	NO		
C9856		SI	SI

Finally, Figure 5 lists the set of queries actually applied to build up the final matrix of textual data, which has been used as the database for the econometric estimation of the Authority's decision model, performed with the SAS statistical software program.

BUYER POWER

"potere contrattuale LAG6 domanda" or "domanda LAG20 potere contrattuale"
"potere contrattuale LAG6 fornitor?" or "fornitor? LAG20 potere contrattuale"
"potere contrattuale LAG6 acquirent?" or "acquirent? LAG20 potere contrattuale"
"potere contrattuale LAG6 client?" or "client? LAG20 potere contrattuale"
"potere contrattuale LAG6 utilizzator?" or "utilizzatore? LAG20 potere contrattuale"
"potere negoziale"
"potere di mercato LAG6 domanda" or "domanda LAG20 potere di mercato"
"potere di mercato LAG6 fornitor?" or "fornitor? LAG20 potere di mercato"
"potere di mercato LAG6 acquirent?" or "acquirent? LAG20 potere di mercato"
"potere di mercato LAG6 client?" or "client? LAG20 potere di mercato"
"potere di mercato LAG6 utilizzator?" or "utilizzatore? LAG20 potere di mercato"

ENTRY BARRIERS

"non LAG4 esist* LAG6 barriere"
"inesisten* LAG2 barriere"
"non LAG4 presen* LAG3 barriere"
"non LAG1 sussist* LAG3 barriere"
"non LAG2 caratteriz* LAG4 barriere"
"non LAG2 riscontr* LAG4 barriere"
"non vi sono LAG2 barriere"
"Assenza LAG2 barriere"
"scars* LAG2 barriere"
"privo LAG2 barriere"
"senza LAG1 barriere"
"bass? LAG2 barriere"
"mancanza LAG2 barriere"
"barriere LAG4 mercato LAG5 assenti"
"rendere elevate le barriere all' entrata sono la reputazione necessaria per operare"
"autostradale presenta rilevanti barriere all' entrata di carattere amministrativo"
"nonché dalla presenza di LAG1 barriere all' entrata"
"barriere all' entrata di tipo regolamentare sono ancora consistenti"
"esistono ancora consistenti barriere all' entrata di tipo istituzionale"
"aerei civili è contraddistinta da barriere all' entrata di tipo tecnologico"
"caratterizzata da stringenti barriere di tipo regolamentativo"
"caratterizzato dall' esistenza di barriere all' entrata"
"barriere all' entrata in questo settore sono LAG1 elevate"
"barriere all' entrata nel mercato sono costituite dal regime"
"principali barriere all' entrata sul mercato consistono"
"ciascuno caratterizzato da rilevanti barriere istituzionali e tecnologiche"
"contraddistinta dalla presenza di barriere all' entrata"
"le principali barriere all' entrata sono costituite"
"Tenendo presente anche le forti barriere all' ingresso"
"esistenza di barriere all' entrata di tipo tecnologico"

VERTICAL EFFECTS
"operazione LAG10 integrazione LAG2 verticale" and not "operazione in esame non rivela aspetti di integrazione"
"operazione comunicata non avrà effetti di integrazione"
"non si verifica alcun fenomeno di integrazione"
"ritiene che l' operazione non determini effetti di integrazione"
"concentrazione LAG10 integrazione LAG2 verticale"
"gli effetti verticali LAG2 riguardano"
"operazione presenta effetti verticali"
"operazione in esame produce effetti verticali nel mercato del calcestruzzo"
"integrazione verticale risultante"
"determina esclusivamente effetti verticali di portata limitata"
"operazione in esame determina prevalentemente effetti verticali"
"gli effetti verticali dell' operazione in esame riguardano l' integrazione a monte"
"per quanto attiene gli effetti verticali derivanti dalla operazione in esame"
"determina effetti verticali" and not "non determina effetti verticali"
"concorrenti LAG1 integrati"
"Presenza LAG4 operatori verticalmente integrati"
"distributori LAG3 integrati"
"produttori LAG5 integrati" and not "non integrati"
"operatori LAG4 integrati" and not "operatori non LAG4 integrati"
"integrazione verticale LAG2 in atto nel settore"
"integrazione verticale delle società petrolifere"
"mercato LAG6 elevat? LAG2 integrazione verticale"
"settore LAG8 elevat? LAG2 integrazione verticale"
"elevat? LAG2 integrazione verticale" and not "non LAG2 elevat? LAG2 integrazione verticale"
"gruppi cartari integrati""gruppi integrati a monte""integrati in linea verticale""integrati nei diversi mercati dell' editoria"
COORDINATED EFFECTS
"non si può escludere LAG10 posizione dominante collettiva"
"comportamenti collusivi"
"effetti coordinati"

FIGURE 5: Queries used to retrieve topics of interest and classify documents