European Accounting Review
Vol. 15, No. 2, 181–218, 2006

Disclosure and Determinants Studies: An Extension Using the Divisive Clustering Method (DIV)

MARIE CHAVENT[*], YUAN DING[**,†], LINGHUI FU[‡], HERVÉ STOLOWY[**] AND HUIWEN WANG[‡]

[*]*MAB, University of Bordeaux I, France,* [**]*HEC School of Management, Paris, France*
[†]*Regular Visiting Professor of Accounting, China–Europe International Business School (CEIBS), Shanghai, P.R. China and* [‡]*Beihang University, Beijing, P.R. China*

ABSTRACT Past accounting research contains an extensive range of disclosure and determinants studies. But these studies have one major methodological drawback: the disclosure analysis is often restricted to determination of the disclosure index, that is, the sum of disclosed items, weighted or unweighted. The disclosure profile (which reflects the structure of published information) is generally not part of the research design. The objective of this paper is to introduce a divisive (descendant) clustering method, which splits the sample into homogeneous sub-groups corresponding to disclosure patterns (or profiles), for clearer determination of the financial characteristics of each group. This methodology is illustrated by a study of disclosure on provisions by large French firms. The results show that the disclosure pattern is related to provision intensity, size, leverage and market expectation, but not to profit, return and industry. This new research method is a valuable complementary tool for expanding on disclosure and determinants studies, moving from disclosure levels to disclosure patterns.

1. Introduction

Accounting disclosure and determinants analysis is a major issue in accounting research (see the abundant literature reviewed in the next section – see also Appendix A). Researchers try to answer two major questions: (1) What attitude do firms take towards accounting disclosure, either general or specific (e.g. disclosure on business segments, R&D activities, environmental projects, social

Correspondence Address: Hervé Stolowy, HEC School of Management, Paris, Department of Accounting and Management Control, 1, rue de la Liberation, 78351 Jouy-en-Josas, France. E-mail: stolowy@hec.fr

0963-8180 Print/1468-4497 Online/06/020181–38
DOI: 10.1080/09638180500253092
Published by Routledge Journals, Taylor & Francis, on behalf of the EAA.

responsibility, etc.)?; (2) Why do some firms disclose more (or less) information than others? The first research question leads to what is known as 'disclosure level evaluation', and the second, 'disclosure determinants analysis'.

The most common method used in the literature consists of calculating a firm-based disclosure score and then running a multivariate linear regression, with this score as dependent variable, and various (financial or other) characteristics of the related firm (firm size, leverage, financial performance, industry, listing status, audit firm size, etc.) as independent variables.

However, there are several limitations to this methodological approach: (1) The disclosure index is often determined by totaling several items. These items can be weighted or unweighted; while many papers raise this issue, there is no one dominant practice; (2) The more independent variables the model contains, the more interesting it will be and the easier it will be to find explanations for a particular disclosure behavior. Including too many variables, however, may create a multicollinearity difficulty (Wallace *et al.*, 1994). Several solutions have been proposed, including the regression of several separate models based on a selection of independent variables, or a factor analysis of all independent variables (Bah and Dumontier, 2001); (3) The form of the relationship between dependent and independent variables is not always known (Lang and Lundholm, 1993; Cooke, 1998). A 'classical' linear regression is not always the most suitable tool, and rank regression may appear an appropriate solution.

These three limitations have been discussed in the literature in varying levels of detail. Yet by aggregating the different disclosure items into one disclosure index, all the existing studies retain only the quantity of disclosure (or the disclosure level). The firms' structure of published information, which we call 'disclosure pattern' or 'disclosure profile', is totally lost in the analysis, although the information is present in the data collected.

The main objective of this paper is thus to propose a new methodological approach to analyze firms' disclosure practices and their determinants. This method is not put forward as a replacement for the traditional way of analyzing the disclosure level and its determinants as described above, but as a complementary approach. We introduce a divisive clustering method (DIV) analysis to understand the different disclosure patterns. Under this method, we split the sample into homogeneous sub-groups corresponding to disclosure patterns (or profiles), for clearer determination of each group's financial characteristics. We can then explain the different disclosure profiles.

This method is illustrated with data concerning the disclosure of information on provisions, collected from the 2001 annual reports of French listed industrial and commercial firms belonging to the SBF 120 index.

We chose the topic of provisions against the background of the successive financial and accounting scandals (Enron, WorldCom, Xerox, AOL, HealthSouth, Ahold and others) that have hit the world since Fall 2001. These affairs have brought accounting manipulation under the spotlight of particular attention in both the business and academic worlds. One of the main measurement instruments for

accounting manipulation is the level of discretionary accruals (Jones, 1991). Provisions make up a large part of discretionary accruals, and this leads us to believe that in the current context, provisions are a highly relevant field for exploring the earnings-adjustment mechanisms used by firms. We also believe that examining disclosure on provisions will give users broader access to information on the accounting practices and choices adopted by firms with regard to provisions.

To the best of our knowledge, disclosure on provisions has never been studied in this way before, especially in France. French firms are allowed to record two categories of provisions. The first category, 'provisions for depreciation' (corresponding to 'valuation allowances' in US accounting), reflects specific potential losses on an asset, for instance, provisions for bad debts, provisions on long-term investments or provisions on inventories. In French accounting, these provisions are deducted from the gross value of assets. The second category is called 'provisions for risks and expenses'. These provisions are established to cover certain general risks or potential losses not linked directly to assets: for example, provisions for restructuring, provisions for foreign exchange losses, provisions for litigation, etc.

In our view, although firms can adjust their earnings by overestimating or underestimating their provisions for bad debts or provisions on inventories, the first category of provisions offers less 'room for maneuver' than provisions for risks and expenses, because it is closely linked to the firm's operating and investing activities. For this reason, our study concentrates solely on the second category – provisions for risks and expenses ('provisions' in the rest of this paper). We are particularly interested in why firms disclose their policies on provision for risks and expenses and how this disclosure level is associated with firms' financial characteristics. This study is made possible because of the flexibility allowed by French regulations in terms of content of the information disclosed.

Although the results of the classical linear regression are weak, they are slightly improved using rank regression. The most interesting results, however, are produced by the DIV analysis. We find that disclosure pattern is related to provision intensity, size, leverage and market expectation, but not to profit, return and industry. We also find that the group which discloses the greatest amount of information has the greatest proportion of provisions, size, leverage and market expectation.

The rest of the paper proceeds as follows. The next section provides a literature review on voluntary disclosure, disclosure adequacy and their determinants, and discusses the methodological issues. Section 3 lays out the hypotheses. Section 4 explains our sample and data collection. Section 5 discusses our research design. Section 6 presents the empirical findings and Section 7 concludes the paper.

2. Literature Review and Methodological Issues

Principle of Disclosure Studies

Healy and Palepu (2001), and a discussion by Core (2001), provide a broad overview of the empirical disclosure literature. More specifically, many researchers

have taken an interest in the corporate characteristics that could predict a firm's disclosure level. We summarize the abundant literature in Appendix A; the reader can also refer to Ahmed and Courtis' (1999) meta-analysis of 29 disclosure studies. The fact is that many firms exceed the disclosure requirements and provide additional information not specifically required by the existing law or accounting standards (Dumontier and Raffournier, 1999). This increasing transparency is supposed to help reduce the firm's agency and political costs. For example, in his study, Raffournier (1995) tries to explain the voluntary disclosure level of Swiss firms by their size, share listing, profitability, ownership structure, use of external financing, size of auditing firm, internationality and industry type. In a literature review on firms' voluntary disclosure decisions, Healy and Palepu (2001) analyze managers' reporting and disclosure decisions in a capital markets setting. They argue that six forces affect managers' disclosure decisions for capital market reasons: capital market transactions, corporate control contests, stock compensation, litigation, proprietary costs and management talent signaling.

Not all the papers concentrate on a general disclosure level. Several analyze more specific types of information: interim reporting (Leftwich *et al.*, 1981), segment information (Mitchell *et al.*, 1995; Aitken *et al.*, 1997; Emmanuel and Garrod, 2004; Leung and Horwitz, 2004; Prencipe, 2004), communication on R&D (Entwistle, 1999) and ratio disclosure (Watson *et al.*, 2002). The link between corporate disclosure policy and analyst behavior has also been investigated (Lang and Lundholm, 1996) as well as the relationship between disclosure level and the cost of equity capital (Botosan, 1997; Botosan and Plumlee, 2002).

Overview of Methodological Issues

As mentioned in the Introduction, several methodological issues have been raised, often independently, by the numerous researchers who have embarked on disclosure studies. Figure 1 summarizes the various issues identified in the past literature.

Determination of the Disclosure Index

The vast majority of disclosure studies adopt an item-based approach using a dichotomous procedure in which an item scores one if it is disclosed and zero otherwise. A few articles use a different approach: for example, number of words used to describe an item disclosed (Copeland and Fredericks, 1968) or a content analysis based on the number of sentences (Entwistle, 1999; Williams, 1999).

In the most common approach, the concept of 'disclosure index' was first used by Buzby (1975, p. 27) and Stanga (1976, p. 48) and formalized by Cooke (1989a, 1989b). We can summarize the determination of the index as follows:

$$\text{Index} = \text{Actual disclosure/Total possible disclosure} = \sum_{i=1}^{m} d_i \Big/ \sum_{i=1}^{n} d_i$$

where $d = 1$ if item d_i is disclosed and 0 if item d_i is not disclosed, $m =$ number

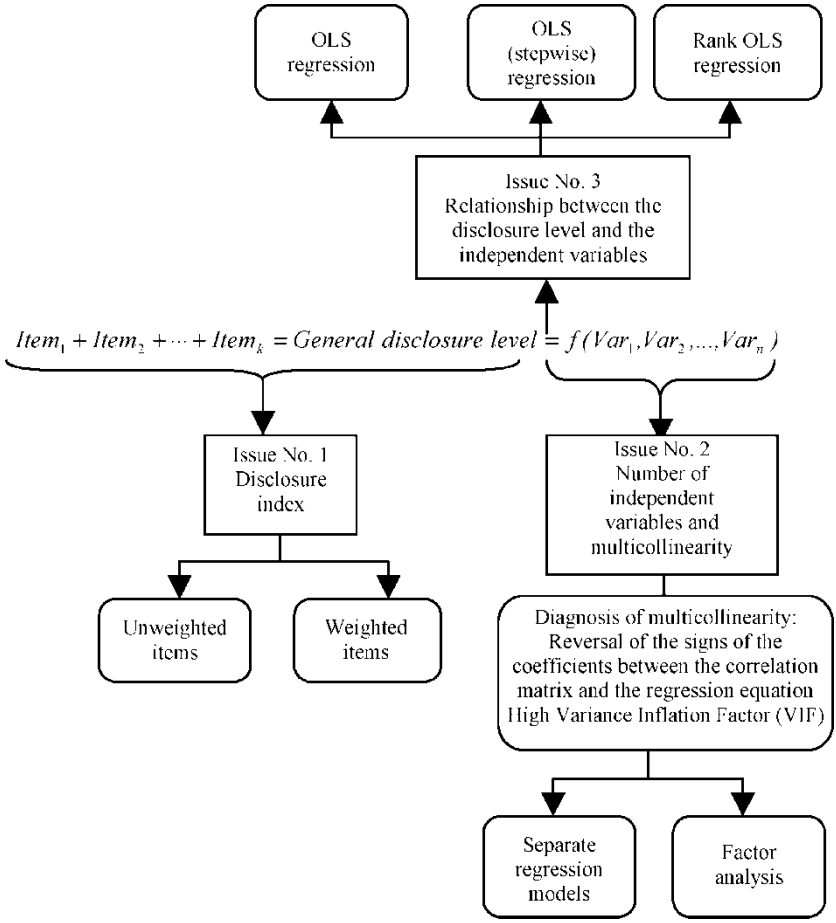

Figure 1. Overview of methodological issues and solutions

of items disclosed and n = maximum number of disclosure items possible. The index is a ratio comparing the actual level of disclosure and the possible level (thus not penalizing the firm for non-disclosure of irrelevant items).

Although there is a general consensus on the determination of this index (see Appendix A for the numerous studies referring to the Cooke index), a debate has arisen concerning item weighting. Cooke (1989b, 1991, 1992, 1993), and many other authors (e.g. Tai *et al.*, 1990; Ahmed and Nicholls, 1994; Hossain *et al.*, 1994, 1995; Wallace *et al.*, 1994; Chen and Jaggi, 2000; Archambault and Archambault, 2003), are in favor of unweighted items, implying that each item is of equal importance. The major argument is that 'one class of user will attach different weights to an item . . . than another class' and that 'the subjective weights of user groups will average each other out' (Cooke, 1989b, p. 115).

Conversely, other authors prefer to apply weighting to the different items. As Ahmed and Nicholls (1994, p. 68) point out, the weighting factors may be predetermined subjectively (Cerf, 1961; Singhvi and Desai, 1971) or taken from prior studies (Barrett, 1977). Finally, some authors stress that certain items are more important to users than others, and send a list of items (questionnaire) to a sample of users, asking them to evaluate the importance of each item (e.g. Buzby, 1975; Stanga, 1976; Firth, 1979; McNally *et al.*, 1982; Giner, 1997).

Marston and Shrives (1991) provide a review of the literature that has made use of the disclosure index as a measurement technique, while Coy *et al.* (1993, p. 123) compare the index construction methods used by different researchers. Finally, Ahmed and Courtis (1999, p. 36) write that the approach based on unweighted items 'has become the norm in annual report studies' because it reduces subjectivity.

Independent Variables and Multicollinearity

To enhance explanations of the disclosure index, the researcher may be tempted to increase the number of independent variables, that is, determinants. But this decision may lead to a higher potential for collinearity between the variables (Singhvi and Desai, 1971; Moore and Buzby, 1972), which is identified by running a correlation matrix. The reversal of the signs of the coefficients on some variables between the correlation matrix and regression equations is another symptom of collinearity (Wallace *et al.*, 1994, p. 49). Collinearity can also be diagnosed by evaluating the variance inflation factor (VIF) for each variable (Patton and Zelenka, 1997; Owusu-Ansah, 1998; Ho and Wong, 2001; Chau and Gray, 2002; Haniffa and Cooke, 2002). The VIF measures the degree to which each explanatory variable is explained by the other explanatory variables. Traditionally, collinearity is not considered to be a problem when the VIF does not exceed 10 (Neter *et al.*, 1983).

Several solutions have been put forward in past studies to solve this multicollinearity issue. First, different regression models are run, each routine using only one of the independent variables identified as generating a multicollinearity problem, in particular size variables (Cooke, 1989a, 1989b, 1991; Ahmed and Nicholls, 1994; Depoers, 2000).

Another solution consists of factoring the collinear independent variables and using the principal factors as regressors (Cooke, 1992; Eng and Mak, 2003).

Relationship between Dependent and Independent Variables

The multiple linear regression method has been used extensively to link the disclosure level to the financial (size, leverage, profitability, etc.) and non-financial (industry, listing status, audit firm size, etc.) variables. However, this method can be applied only if several conditions are met:

- The variables have a normal distribution.
- The error term has a normal distribution with a mean of zero. The variance of the error term is constant across cases and independent of the variables in the model.

- The value of the error term for a given case is independent of the values of the variables in the model and the values of the error term for other cases.
- There is no multicollinearity among independent variables.

Several tests can be conducted to verify the fundamental assumption of normality: skewness, kurtosis and Kolmogorov–Smirnov. But even if the results of these tests are positive, a major difficulty has been identified: the 'theoretically correct form of the relation between [the disclosure index] and the independent firm variables is not known' (Lang and Lundholm, 1993, p. 261). In other words, the linearity is only an assumption. These authors suggest using rank regressions to analyze data, citing Iman and Conover (1979, p. 508) who write that 'rank regressions are quite powerful when the relations are nonlinear but monotonic'. In practice, it is necessary to transform both independent and dependent variables into ranks before applying the OLS regression (Wallace *et al.*, 1994, p. 47; Wallace and Naser, 1995).

Following in the footsteps of Lang and Lundholm (1993), a few authors have applied the rank regression in the context of disclosure studies (Wallace *et al.*, 1994, p. 47; Wallace and Naser, 1995; Owusu-Ansah, 1998). Some have even used both procedures (with unranked and ranked data) in order to compare the results (Wallace and Naser, 1995).

Independently of the issue of the nature of the relationship between dependent and independent variables, several authors have used the stepwise procedure (Malone *et al.*, 1993; Ahmed and Nicholls, 1994; Raffournier, 1995; Giner, 1997; Depoers, 2000). As explained by Cooke (1991), one way to specify the regression model correctly is to adopt this stepwise procedure, which adds variables to the model to maximize R^2 or equivalently minimize the error sum of squares. This approach is useful in determining which variables should be included in the model.

Given all the methodological issues surrounding regression analysis, Cooke (1998) reviews a number of transformations of data, including the rank regression.

Limitation of the Disclosure Index

One major limitation of linear regression lies in its application to the total disclosure index (addition of items) rather than to the pattern (profile) of items. This means the regression method cannot reflect the structure of disclosure, although divergences between firms in their types of disclosure are to be expected. And these divergences are not reflected in the research design, because the items are combined for incorporation into the disclosure index. The regression thus reduces the richness of the study. In other words, the often highly time-consuming work done by the researcher to determine the disclosure index is 'thrown away' at the regression stage.

3. Hypotheses

After surveying the various methodological issues inherent to disclosure studies, we will introduce a new method in Section 5, illustrated with a study of disclosure on provisions in France. A provision, defined as a liability of uncertain timing or amount (IASC, 1998), is calculated to cover general risks or potential losses not linked directly to assets. For example, French firms record provisions for restructuring, foreign exchange losses, litigation, etc. Pension liabilities are also recorded as provisions in France.

Our hypotheses are of mixed origin, being adapted from general disclosure studies while integrating the specialized topic of provisions. They concern the determinants of firms' provision disclosure level and cover the following characteristics of the firm: provision intensity, size, leverage, profitability, performance, market expectation for the firm's future growth and industry type.

Provision Intensity

The positive link between the relative proportion of provisions and their disclosure level is a rational consequence of application of the materiality principle. When provisions are equivalent to a high percentage of total assets, they become a major factor in evaluating the firm's risk level.

Our first hypothesis is:

H1: The extent of disclosure on provisions is positively related to the firm's provision intensity.

Size

The literature is in agreement on the positive relationship between the firm's size and its information disclosure level. There are at least three reasons for this link. First of all, large firms are more willing to disclose information to reduce their political costs, since their higher visibility can easily lead to more litigation and governmental intervention (Watts and Zimmerman, 1978; Bujaki and Richardson, 1997). Secondly, thanks to their more developed internal reporting system, the costs associated with a higher disclosure level are lower for large firms. Thirdly, smaller firms are more likely to hide crucial information because of their competitive disadvantage within their industry (Singhvi and Desai, 1971; Firth, 1979). Appendix A provides numerous examples of disclosure studies which identify size as a significant determinant (e.g. Stanga, 1976; McNally *et al.*, 1982; Chow and Wong-Boren, 1987; Tai *et al.*, 1990; Hossain *et al.*, 1994; Meek *et al.*, 1995; Marston and Robson, 1997; Depoers, 2000).

Our second hypothesis is thus the following:

H2: The extent of disclosure on provisions is positively related to the firm's size.

Leverage

Corporate information disclosure is often considered as an instrument to reduce the monitoring costs for creditors. We can thus expect a positive link between a firm's disclosure level and its indebtedness, since in the event of high leverage, creditors will urge the firm to disclose more information to help them handle their own credit risk (Hossain *et al.*, 1994). For example, some studies show that diversified firms obtaining long-term capital externally were more likely to disclose segmental financial data voluntarily (Salamon and Dhaliwal, 1980). We expect this relationship to be more visible in our case, since provisions are widely suspected of being an earnings management tool. The disclosure level is therefore very precious as information to help creditors correctly evaluate the risk.

Our third hypothesis is:

H3: The extent of disclosure on provisions is positively related to the firm's leverage.

Profit and Return

Profit and return have also been considered as relevant explanatory variables for the disclosure level (Singhvi and Desai, 1971; Wallace and Naser, 1995). We thus expect a positive relationship between a firm's provision disclosure level and its profitability (Giner, 1997). In their paper, Singhvi and Desai (1971) propose that when the rate of return is high, managers are motivated to disclose detailed information in order to support the continuance of their positions and remuneration. Conversely, when the rate of return is low, they may disclose less information in order to conceal the reasons for losses or declining profits.

We thus arrive at the two following hypotheses:

H4: The extent of disclosure on provisions is positively related to the firm's level of profit.

H5: The extent of disclosure on provisions is positively related to the firm's rate of return.

Market Expectation for the Firm's Future Growth

One of the major roles of financial information disclosure is to reduce agency conflicts and minimize the firm's capital cost by resolving the information asymmetry problem between the principal (shareholders) and the agent (managers) (Jensen and Meckling, 1976). Therefore, a higher provision disclosure level should be associated with a better market expectation for the firm's future growth.

Our sixth hypothesis is:

H6: The extent of disclosure on provisions is positively related to the market expectation for the firm's future growth.

Industry Type

Several studies have highlighted a relationship between the disclosure level and the industry sector (Cooke, 1992; Raffournier, 1995). Conversely, other studies have found no differences in disclosure level between industries (Watson *et al.*, 2002). Although the evidence is inconclusive, we believe it is appropriate to assume there is no link when testing whether disclosure varies between industries (as in Watson *et al.*, 2002).

Our seventh hypothesis is:

H7: The extent of disclosure on provisions is not related to the firm's industry.

4. Sample and Data Collection

Sample

Our statistical survey concerning the disclosure of provisions is based on a sample of large French groups included in the SBF 120 stock index for the year 2001. This choice allowed us to work on a sufficiently large sample, and to carry out a relevant survey covering major sectors of the French economy. It is important to note that only industrial, commercial and service sectors are included in the survey; financial firms, that is, banks, insurance and such sectors as leasing companies are excluded because they use sector-specific accounting principles and their financial statements are not comparable to those of other economic sectors. Our research covers only one year because firms' disclosure policies appear to remain relatively constant over time (Botosan, 1997, p. 327).

A few more companies are excluded from the sample due to the absence of published consolidated financial statements or notes to the financial statements, or because there were no associated data in the Worldscope database. After these adjustments, 100 industrial and commercial French listed companies belonging to the SBF 120 index are included in our study (see Table 1 and the list of the firms in Appendix B).

Data Collection and Explanatory Variables

To explore the provision disclosure pattern and level, we analyzed the 2001 annual reports of the 100 firms included in our sample.

Table 1. Sample construction

Firms belonging to the SBF 120 index	120
Banks, insurance and real estate companies	−15
Firms not disclosing consolidated financial statements or notes, or Worldscope data not available (Arcelor, Autoroutes du Sud de la France, Bouygues Offshore, STMicroelectronics and Transiciel)	−5
Final sample studied	100

We find three categories of studies in the literature:

- Voluntary disclosure: the researcher examines the link between voluntary publication of information and certain determinants. This is a classic, 'natural' research question, and these studies seem to represent the majority of past research (see Appendix A).
- Mandatory disclosure: studying this aspect may appear less logical; after all, if publication of certain information is mandatory, how can there be differences between firms' disclosures? But in fact, even when disclosures are mandatory, researchers have found that firms still have some flexibility in the way they report the information. This is referred to as 'disclosure extensiveness of each item of mandatory disclosure' (Chen and Jaggi, 2000). A number of such studies also exist (see Appendix A).
- Mandatory and voluntary disclosure: numerous studies (see Appendix A) cover both types of item (e.g. Cooke, 1993). Our paper relates to this third category.

The disclosure index is based on the addition of several items. When a study covers both voluntary and mandatory disclosures (remembering that mandatory disclosures may have a degree of flexibility in the content – see above), there is no predefined list of items, so it is necessary to define a list which will be applied to all the companies surveyed in order to compute the disclosure index. This list of items is often determined on the basis of past literature and regulations (e.g. Adhikari and Tondkar, 1992, p. 87). In the present case, we are not aware of any past literature relating to disclosure on provisions. The only relevant French standard is Regulation 99-02 on Consolidation (X, 1999). In respect of provisions (for risks and expenses), this regulation states (§ 424), without any further stipulation, that the notes must include an 'analysis, with comments, of the main balances and movements'. The wording leaves room for significant flexibility in terms of content of the information disclosed. This being the case, we generated a preliminary list of items on the basis of a sample of annual reports, and finalized this list based on a study of all the annual reports.

Fourteen information items were identified (see the list in Appendix C). Of these 14, given the flexibility allowed by the accounting regulation, only one can be considered as mandatory: the Year1/Year0 comparison (item 8). Regulation 99-02 (X, 1999) states that the information disclosed in the notes must concern the current and preceding year. Although we could have applied the research design to all 14 items, we decided, for the sake of clarity, to drop item 8 from the list.

The items studied belong to two categories:

- Items present or absent in the published information (e.g. item 1: inclusion in the notes of certain information on provisions). In this case, the item is coded 1 if it is disclosed, 0 otherwise.
- Items subject to a condition (e.g. item 14: amount of 'other provisions' lower than the average for the sample). In this case, the item is coded 1 if the condition is met, 0 otherwise.

The coexistence of items scored on a 'present/absent' basis and items subject to a condition has already been observed (Barrett, 1976; Marston and Robson, 1997) and raises no specific methodological issue.

The dummy variables corresponding to the 14 items studied reflect the firm's effort in terms of transparency and accuracy of information. These dummy variables not only define the position of each firm regarding its provision disclosure structure, but also give us an aggregated vision of firms' general provision disclosure level.

In comparison to other disclosure studies (see Appendix A), the number of items is limited. But our survey of the literature provides some examples of studies carried out on a limited number of items: nine items (Prencipe, 2004), 17 items (Marston and Robson, 1997). This small number of items results from the scope of the study, which concentrates on a single topic, provisions, and is consistent with past literature (e.g. Prencipe, 2004, in the field of segment information).

We then extract the financial data for our sample from the annual reports or from the Thomson Analytics' Worldscope database. The analyzed variables, which constitute proxies for our hypotheses, are described in Table 2.

Past research has used various classifications to group firms by category: conglomerate, manufacturing, services, trading (Cooke, 1991): (1) metals, building materials and construction; (2) engineering; (3) consumer goods and services; (4) oil, chemicals and mining (Meek *et al.*, 1995); manufacturing, non-manufacturing (Cooke, 1992; Raffournier, 1995); basic industry, manufacturing industry, service industry (Giner, 1997): (1) electronics and technology; (2) publishing and printing; (3) food and beverage; (4) shipping and transportation (Chau and Gray, 2002). We have adopted a more detailed division into nine sectors: automobile, building, consumer goods, energy, food, health, industry, media and technology. This industry

Table 2. Explanatory variables

Hypotheses	Name of the variable	Source	Explanation (if necessary)
H1 Provision intensity	Provision intensity	Computation from the annual report	Provisions/Total assets
H2 Size	Sales	Annual report	–
H3 Leverage	Total Debt Pct Common Equity	Worldscope	(Long-term debt + Short-term debt & Current portion of long-term debt)/ Common equity * 100
H4 Profit	Profit	Derived from Worldscope	Income/Sales
H5 Return	Return On Equity Total	Worldscope	–
H6 Market expectation	Dividend Yield Current	Worldscope	–
H7 Industry	Industry (Dow Jones Classification)	Worldscope	–

grouping is based on the Dow Jones Classification found in the Worldscope database, simplified to nine sectors.

5. Research Design

In a preceding section, we discussed the various methodological issues surrounding disclosure studies. In order to avoid the problems referred to, we adopt a research design comprising two steps (see Figure 2). Firstly, we run a linear regression, with unranked then ranked data in order to test the relationship between the disclosure level, used as the dependent measure, and the different determinants. Secondly, we use the DIV method to identify the information items that determine a firm's provision disclosure pattern, and classify firms in different clusters according to pattern. On the basis of this clustering, we will test our hypotheses by analyzing the differences between our clusters' financial characteristics based on their provision disclosure pattern.

Disclosure Index

We decided to adopt an unweighted index, treating all items equally, given that there is no specific set of users of information on provisions (Cooke, 1989b). The index is the ratio of the total score awarded to a company to the maximum possible score that company could obtain.

Linear Regression

The limits of linear regression have been discussed above. We will use two of the methods mentioned: the OLS regression and the rank OLS regression, and check for multicollinearity by computing the VIFs.

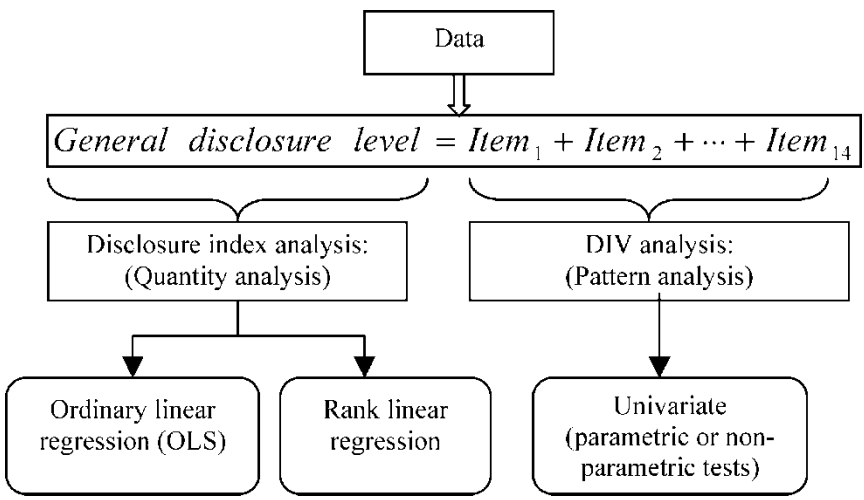

Figure 2. Research design

DIV Analysis

The regression method cannot reflect the structure of provision disclosure. We can expect divergences between firms in the types of provision information. For example, some companies prefer to disclose information on the treatment of provisions, while others like to inform users about a particular type of provision (e.g. for restructuring). This structural information cannot be reflected in the general provision disclosure level, and the regression will therefore reduce the richness of the study.

To deal with this difficulty, we introduce the DIV method, a divisive clustering method. It simultaneously defines a hierarchy of a set of objects and a monothetic definition of each cluster in the hierarchy (Chavent, 1998). It is similar to a segmentation method, and starts with all objects in one group, dividing each group successively into smaller ones. DIV is also a monothetic clustering method. A cluster is called monothetic if a conjunction of logical properties is both necessary and sufficient for membership of the cluster (Sneath and Sokal, 1973). At each stage, bipartition is performed by a single variable and a specified value of the variable in the monothetic clustering. DIV analysis is a descendant hierarchical clustering method (Chavent *et al.*, 1999). It is very different from the hierarchical clustering analysis found in statistical software (e.g. SPSS) and already used in past research (Stolowy and Tenenhaus, 1998; Sucher *et al.*, 1999). While this second method is an *ascendant* (or agglomerative) hierarchical clustering technique, DIV is *descendant*, as already stated. Consequently, the major advantage of DIV analysis is that it explains the origin of the clustering by determining which items separate the companies into groups.

DIV analysis is also different from recursive partitioning, although the two analyses are similar in that they both produce a decision tree as output. (For the DIV method, this decision tree is strictly speaking a dendrogram of a hierarchy.) Recursive partitioning, which has often been used in scientific research, for example, in chemistry (Downs and Barnard, 2002), has rarely been applied in accounting research, with the notable exception in auditing of the work of Cormier *et al.* (1995). Recursive partitioning classification is a supervised technique (hence the term 'classification') while DIV is a non-supervised technique (hence the term 'clustering'). The major difference between the two methods lies in the following: in recursive partitioning, the researcher has identified the partition beforehand and wants to explain it. This is an explanatory method which is also closely related to discriminant analysis. Conversely, the DIV method is applied without a pre-identified partition and aims to develop a clustering pattern with homogeneous groups.

Cormier *et al.* (1995), for example, want to explain a binary variable (failing firms versus non-failing firms), which constitutes a partition into two classes. They split the group using a question which includes a threshold, then split one of the sub-groups on the basis of another question, and so on. In other words, the classification (not 'clustering') is supervised by the existence of the

binary variable. In DIV, as we will see below, clustering is automatic because the data-set is first divided into two daughter groups that are as different as possible. Each of these groups is then divided again, and so on (Urban, 2004).

DIV was developed in the framework of symbolic data analysis. It is also applied in standard data analysis (see Appendix D for details). DIV analysis was carried out using the 'SODAS' software tool, developed jointly by 17 European teams (sponsored by EUROSTAT) and available free of charge. SODAS facilitates the use of analysis techniques for numeric or symbolic data, and can be used in particular for data with a complex structure, to provide better explanations of statistical results, and to represent, manipulate or better analyze concepts and metadata.

As mentioned above, disclosure level study results in the existing literature are often limited to the aggregated final disclosure score. The DIV method makes it possible to break through this limitation. Here, we will explore the provision disclosure structure of each firm. This clustering method will classify the firm sample into several groups according to their provision disclosure patterns (which information they emphasize and which information they play down). The criteria used for classification will also be crucial: the presence or absence of one specific information item can predict the disclosure pattern of a firm.

Determinants Analysis between Different Groups

The usual way of analyzing the link between a firm's provision disclosure level and its financial characteristics is the linear regression method. For the reasons described above, we will add the statistical method, measuring the differences in financial characteristics between groups of firms classified according to their provision disclosure patterns. We will use either parametric or non-parametric tests depending on the distribution patterns of the financial characteristics.

6. Results

Regression Model

The regression model below is applied, first with unranked data, then with ranked data (following Wallace and Naser, 1995):

$$\text{Index} = \alpha_0 + \alpha_1 \text{prov\_ass} + \alpha_2 \text{size} + \alpha_3 \text{leverage} + \alpha_4 \text{profit} + \alpha_5 \text{return} + \alpha_6 \text{market} + \sum_{l=1}^{8} \alpha_{7,l} \text{industry}_l + \varepsilon$$

where:

Index	= Disclosure index
α_s	= Regression parameters
prov_ass	= Provisions/Total assets
size	= Sales

leverage = Total debt/Equity
profit = Income/Sales
return = Return on equity (ROE)
market = Dividend yield
industry_l = Dummy variable equals 1 if the firm belongs to the following sector:
auto = Automobile
building = Construction
consumer = Consumer goods
energy = Energy
health = Health
industry = Industry
media = Media
techno = Technology
Equals 0 otherwise.
We define the model excluding the 'Food' sector (which has the lowest Provisions/Total assets ratio of the nine sectors). In a linear regression, a categorical variable (i.e. one which can take several values, e.g. the business sector) should be split between a number of dummy variables equal to the number of possible values minus one. The excluded value (here the food industry) serves as a reference for the other dummy variables and this procedure will avoid perfect collinearity.

It should be borne in mind that we are assuming there is information to disclose, and we are interested in how much of it is disclosed, and why. Having said that, if there is no provision recorded in the balance sheet, the research design is not operational. This is related to the definition of the index mentioned in Section 2: actual disclosure/total possible disclosure. If the denominator equals zero, no index can be computed.

Regression with Unranked Data

Table 3, panel A discloses the results of the regression with unranked data. The VIF does not exceed 10 (Neter *et al.*, 1983), which means that multicollinearity is not a real issue. (The correlation matrix [not tabulated] does not disclose high significant correlations.) In this situation, there is no need to implement one of the solutions mentioned in Section 2 (use of different specifications for the same model or factor analysis). The overall model is significant (p-value of $F = 0.035$) and the coefficient of determination (R^2) is 0.240. The adjusted R^2 is 0.115. Disclosure increases with the proportion of provisions over total assets (significant at the 1% level). This result is logical: we would expect a firm to disclose more information if the proportion of provisions is higher. The 'profit' variable is also significant (at the 10% level). All other variables are non-significant.

Table 3. OLS regression

	Panel A – Unranked data Dependent variable: Index			Panel B – Ranked data Dependent variable: Rank of Index		
	b (unstandardized coefficients)	Sig.	VIF	*b* (unstandardized coefficients)	Sig.	VIF
prov_ass	1.099	0.006***	1.300			
size	0.000	0.812	1.420			
leverage	0.000	0.427	1.170			
profit	0.001	0.080*	1.120			
return	0.001	0.363	1.190			
market	0.012	0.359	1.230			
auto	0.116	0.301	1.870	20.669	0.208	1.770
building	0.094	0.339	1.970	6.699	0.655	2.050
consumer	0.012	0.873	3.490	4.731	0.682	3.490
energy	0.103	0.431	1.560	20.316	0.283	1.450
health	0.134	0.292	1.490	16.260	0.408	1.570
industry	−0.003	0.971	3.440	2.499	0.836	3.500
media	−0.039	0.670	2.160	−1.332	0.924	2.240
techno	0.018	0.827	3.240	6.696	0.594	3.380
rank of prov_ass				0.402	0.001***	1.600
rank of size				−0.091	0.475	1.870
rank of leverage				0.193	0.092*	1.490
rank of profit				0.120	0.252	1.260
rank of roe				0.085	0.412	1.250
rank of market				0.076	0.473	1.300
Constant	0.328	0.000		5.370	0.724	
Number of observations	100			100		
F	1.919			2.175		
Prob. > F	0.035**			0.015**		
R^2	0.240			0.264		
Adjusted R^2	0.115			0.143		

*Significant at the 10% level; **significant at the 5% level; ***significant at the 1% level.

As a sensitivity test, we rerun the regression excluding Industry, which does not appear to be a significant explanatory variable (results not tabulated). The *F*-test is significant ($p = 0.0026$) and adjusted $R^2 = 0.1394$. The results are similar.

Regression with Ranked Data

Given that the Index variable does not have a normal distribution (Kolmogorov–Smirnov test) and that there is uncertainty over the nature of the relationship, we transformed both independent and dependent continuous variables into ranks

before applying the OLS regression (with the instruction Transform/Rank cases/ Rank of SPSS). As in Lang and Lundholm (1993), firms with tied ranks were assigned the average of the ranks they would have had if they had not been tied. Results are provided in panel B of Table 3. They are more significant (p-value of $F = 0.015$) and the adjusted $R^2 = 0.143$. While the Provisions/Total assets variable is significant ($p = 0.001$), leverage also appears to be significant (at the 10% level). As a sensitivity test, we rerun the regression excluding Industry. The results remain unchanged (p-value of $F = 0.000$, adjusted $R^2 = 0.179$).

Disclosure Pattern

We voluntarily started with the traditional analysis (regression on unranked then ranked data) and obtained rather poor results. The need for an alternative method is important in this context and will add a valuable extra dimension to the research design.

Based on the disclosed information items, we use the divisive clustering method to classify our sample. The number of clusters can be defined a priori. For the sake of simplification and clarity of results, we chose a three-group clustering. The results of the DIV method are presented in the form of a clustering tree (see Figure 3).

From the clustering according to provision information disclosure, we can observe that the information items 2 (Inclusion in the notes of a description of a particular type of provision) and 6 (Inclusion in the notes concerning pensions liabilities of information on the computation of provisions) play a dominant role in firm classification. (We also ran the DIV clustering with four classes, and item 10 appeared as the third decision node). In other words, the disclosure of item 2

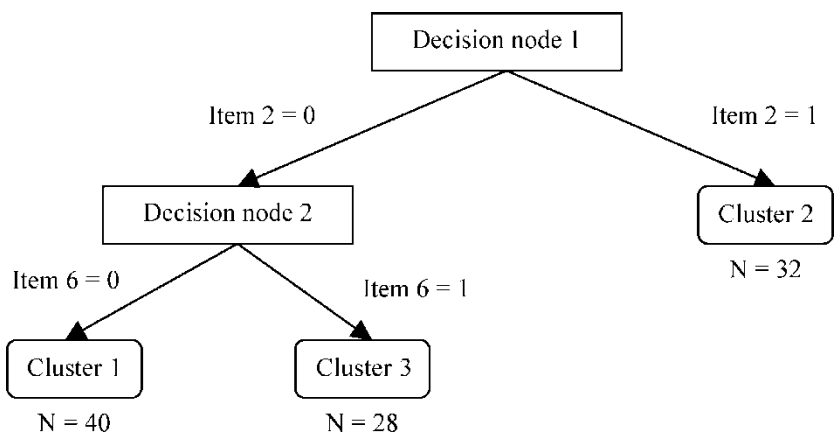

Figure 3. Clustering tree.

will broadly determine the firm's overall disclosure pattern (or profile) on provision information.

Table 4 shows the descriptive statistics of these three firm clusters, stating the percentages of firms in each cluster disclosing the related item. For example, item 1 (Inclusion in the 'accounting principles' part of the notes to the financial statements of a specific note on the treatment of provisions) is disclosed by 20% of firms in cluster 1 (i.e. 8 firms), 90.63% of firms in cluster 2 (29 firms) and 21.43% of firms in cluster 3 (6 firms).

Regarding total disclosure level, Table 4 also shows that firms in cluster 2 disclose the highest quantity of information on their provisions (average disclosure level = 0.649), followed by cluster 3 (0.486) and cluster 1 (0.309). We carry out a one-way ANOVA test with the Scheffe option to check whether the average disclosure level of each cluster was different. The result of this test is highly significant. (An equivalent non-parametric test will be presented in Table 5. It yields similar results.)

By construction, the DIV method has created the clusters in such a way that the presence of the items is different between groups. However, to confirm this idea, we carry out a chi-square test per item. The results are added to Table 4. This test provides evidence that for most of the items, the percentages are different between clusters. Especially for items 2 and 6, which, as seen above, are the decision nodes, the results are highly significant (at the 1% level). However, for some of the items (e.g. No. 13), the results are far from

Table 4. DIV results (descriptive statistics)

Percentages of 1 (item disclosed) within each group	Group 1	Group 2	Group 3	p-Value (χ^2-test)
Item 1	20.00	90.63	21.43	0.000
Item 2	0.00	100.00	0.00	0.000
Item 3	2.50	50.00	0.00	0.000
Item 4	45.00	90.63	82.14	0.000
Item 5	5.00	15.63	32.14	0.011
Item 6	0.00	68.75	100.00	0.000
Item 7	80.00	84.38	71.43	0.461
Item 9	72.50	90.63	85.71	0.115
Item 10	52.50	81.25	57.14	0.032
Item 11	22.50	18.75	42.86	0.079
Item 12	5.00	18.75	21.43	0.102
Item 13	42.50	53.13	46.43	0.667
Item 14	55.00	81.25	71.43	0.054
Number of firms	40	32	28	
Average disclosure level	0.309	0.649	0.486	

Item 8 was dropped because it was published by all the firms in the sample.

Table 5. Kruskal–Wallis tests of differences between the three sub-groups

Variables	Groups	*N*	Mean rank	Chi-square	Asymp. sig.
index	1	40	25.688		
	2	32	79.031		
	3	28	53.339		
	Total	100		61.365	0.000***
prov_ass	1	40	41.600		
	2	32	60.031		
	3	28	52.321		
	Total	100		7.329	0.026**
size	1	40	42.600		
	2	32	56.563		
	3	28	54.857		
	Total	100		4.995	0.082*
leverage	1	40	41.625		
	2	32	58.813		
	3	28	53.679		
	Total	100		6.707	0.035**
profit	1	40	48.200		
	2	32	56.590		
	3	28	46.820		
	Total	100		2.113	0.348
return	1	40	49.025		
	2	32	55.063		
	3	28	47.393		
	Total	100		1.216	0.544
market	1	40	39.388		
	2	32	60.422		
	3	28	55.036		
	Total	100		10.440	0.005***

*Significant at the 10% level; **significant at the 5% level; ***significant at the 1% level.

significant, which shows that the division between the three clusters is not perfect.

Table 4 shows descriptive statistics of these three firm clusters. The patterns appearing in Table 4 are not 'important' per se but are a preliminary step to the implementation of the determinants analysis (see below).

Determinants Analysis between Different Groups

Because the distribution of the explanatory variables is not normal (as shown by the Kolmogorov–Smirnov test for normality), it is necessary to consider using the non-parametric procedures designed to test for the significance of the difference between clusters. These procedures are called non-parametric because they make no assumptions about the parameters of a distribution, nor do they assume

that any particular distribution is being used (Siegel and Castellan, 1988). One popular non-parametric test for independent samples greater than two is the Kruskal–Wallis test. Table 5 shows the mean rank per cluster, for each variable, and the results from the Kruskal–Wallis test for differences between clusters 1, 2 and 3.

Table 5 should be read in conjunction with Table 4. The three clusters determined with the DIV method correspond to three different patterns of disclosure.

The results for the 'index' variable indicate a very significant difference between clusters (at the 1% level). Firms in cluster 2 disclose more information. As already noted, in this context, item 2, identified with the DIV method, will have an impact on the disclosure profile of the firms. In other words, a firm which discloses this item will belong to cluster 2 and will disclose significantly more items than firms in the other clusters.

This result is equivalent to that yielded by the linear regression, in that it explains the quantity of disclosure (firms belonging to cluster 2 disclose the highest number of information items) with determinants: firms in cluster 2 have the highest provision intensity, the largest size, the highest leverage and the highest market expectation.

These conclusions are drawn from the rest of Table 5, which shows that most of our hypotheses are validated regarding the *pattern* (or *profile*) of disclosure. Based on the disclosure pattern, as identified with the DIV method, the firms in the three clusters differ in: the ratio of provisions to total assets (at the 5% level), size (at the 10% level), leverage (at the 5% level) and market expectation (at the 1% level). The differences between clusters show that these determinants, that is, provision intensity, size (to a lesser extent), leverage and market expectation play a major role in a firm's inclusion in one of these clusters, and hence to the nature of the information disclosed.

Table 5 provides evidence of the advantages of the DIV method over the linear regression in our study:

- The first advantage lies in the disclosure pattern identified. Each cluster corresponds to a given pattern of disclosure and the list of items can be read in Table 4. With Table 5, each pattern can be explained with the significant variables. The determinants of cluster 2 are defined above. To conclude on this issue, the determinants can explain not only the general level of disclosure, but also the nature of the items disclosed.
- Whereas the results of the unranked and ranked linear regression were modestly significant, the DIV method has identified three clusters of firms, and the difference in the level of disclosure (the 'index' variable in Table 5) is highly significantly different between those three clusters. In other words, the quantity of disclosure (higher for cluster 2) is better explained with the DIV method.

Limitations and Directions for Future Research

We believe that the DIV method is a valuable extension to the traditional method (linear regression) used in disclosure and determinants studies, but will by no means replace the linear regression, since these two methods concern two important and different aspects of disclosure: the disclosure level for linear regression and the disclosure pattern for the DIV method.

The major limitation arises from the basic principle of the DIV method: the number of clusters is arbitrary, although the creation of more clusters (subgroups) does not really change the general interpretation of the results. In our provision study, given the limited number of items (only 13), the number of clusters cannot be too high. However, if the method was applied to a higher number of items, the number of clusters could be increased (for example, to four or five).

The major opportunity for future research would be to extend this DIV method to other areas of disclosure, whether general or specialized.

7. Conclusion

This study looks at the disclosure pattern for information related to provisions reported by large French listed industrial and commercial firms. The firms' financial characteristics linked to this disclosure pattern are also studied.

By investigating the annual reports for 2001 of 100 French firms belonging to the SBF 120 stock index, we found that the disclosure pattern is associated with provision intensity, size, leverage and market expectation. Our study also shows that firms with the highest score for disclosures have the greatest provision intensity, firm size, leverage and market expectation.

The objective of this paper is to make two contributions to the literature. Firstly, this study uses an innovative statistical approach to analyze firms' disclosure patterns and levels. Secondly, it contains the first empirical exploration in the field of disclosure on provisions and the relationship between a firm's provision disclosure and its financial characteristics, thus enriching the existing literature, especially as regards voluntary disclosure issues.

Acknowledgments

Marie Chavent is a member of the Mathématiques Appliquées de Bordeaux (MAB) (UMR-5466 CNRS). Yuan Ding and Hervé Stolowy are members of the GREGHEC, CNRS Unit, UMR 2959. The authors have benefited from helpful comments by Michel Tenenhaus and workshop participants at the University of Bordeaux IV (CRECCI, France, 2003), the International Conference on Research Methods 'Crossing Frontiers in Quantitative and Qualitative Research Methods' (Lyon, France, 2004) and the EAA Annual Congress (Prague, Czech

Republic, 2004). The authors are particularly grateful to the two anonymous reviewers for their insightful suggestions. Hervé Stolowy has received financial support from HEC School of Management, Paris (project A0426). Yuan Ding and Hervé Stolowy would also like to thank Nicolas Loock and Germain Le Graverend for their able research assistance. Finally, the authors are indebted to Ann Gallon for her much appreciated editorial help.

References

Adhikari, A. and Tondkar, R. H. (1992) Environmental factors influencing accounting disclosure requirements of global stock exchanges, *Journal of International Financial Management and Accounting*, 4(2), pp. 75–105.

Ahmed, K. and Nicholls, D. (1994) The impact of non-financial company characteristics on mandatory disclosure compliance in developing countries: the case of Bangladesh, *The International Journal of Accounting*, 29(1), pp. 62–77.

Ahmed, K. (1996) Disclosure policy choice and corporate characteristics: a study of Bangladesh, *Asia-Pacific Journal of Accounting*, 3(1), pp. 183–203.

Ahmed, K. and Courtis, J. K. (1999) Associations between corporate characteristics and disclosure levels in annual reports: a meta-analysis, *British Accounting Review*, 31, pp. 35–61.

Aitken, M. *et al.* (1997) Determinants of voluntary disclosure of segment information: a re-examination of the role of diversification strategy, *Accounting and Finance*, 37, pp. 89–109.

Archambault, J. J. and Archambault, M. E. (2003) A multinational test of determinants of corporate disclosure, *The International Journal of Accounting*, 38(2), pp. 173–194.

Bah, R. and Dumontier, P. (2001) R&D intensity and corporate financial policy: some international evidence, *Journal of Business Finance and Accounting*, 28(5/6), pp. 671–692.

Barrett, M. E. (1976) Financial reporting practices: disclosure and comprehensiveness in an international setting, *Journal of Accounting Research*, 14(1), pp. 10–26.

Barrett, M. E. (1977) The extent of disclosure in annual reports of large companies in seven countries, *The International Journal of Accounting*, 12(2), pp. 1–25.

Botosan, C. A. (1997) Disclosure level and the cost of equity capital, *The Accounting Review*, 72(3), pp. 323–349.

Botosan, C. A. and Plumlee, M. A. (2002) A re-examination of disclosure level and the expected cost of equity capital, *Journal of Accounting Research*, 40(1), pp. 21–40.

Bujaki, M. and McConomy, B. J. (2002) Corporate governance: factors influencing voluntary disclosure by publicly traded Canadian firms, *Canadian Accounting Perspectives*, 1(2), pp. 105–139.

Bujaki, M. L. and Richardson, A. J. (1997) A citation trail review of the uses of firm size in accounting research, *Journal of Accounting Literature*, 16, pp. 1–27.

Buzby, S. L. (1975) Company size, listed versus unlisted stocks, and the extent of financial disclosure, *Journal of Accounting Research*, 13(1), pp. 16–37.

Cahan, S. F. *et al.* (2005) Global diversification and corporate disclosure, *Journal of International Accounting Research*, 4(1), pp. 73–93.

Cerf, A. R. (1961) *Corporate Reporting and Investment Decisions* (Berkeley: University of California Press).

Chau, G. K. and Gray, S. J. (2002) Ownership structure and corporate voluntary disclosure in Hong Kong and Singapore, *The International Journal of Accounting*, 37(2), pp. 247–265.

Chavent, M. (1998) A monothetic clustering method, *Pattern Recognition Letters*, 19, pp. 989–996.

Chavent, M. *et al.* (1999) Méthodes divisives de classification et segmentation non supervisée: Recherche d'une typologie de la peau humaine saine, *Revue de Statistique Appliquée*, XLVII(4), pp. 87–99.

Chen, C. J. P. and Jaggi, B. (2000) Association between independent non-executive directors, family control and financial disclosures in Hong Kong, *Journal of Accounting and Public Policy*, 19, pp. 285–310.

Chow, C. W. and Wong-Boren, A. (1987) Voluntary financial disclosure by Mexican corporations, *The Accounting Review*, 62(3), pp. 533–541.

Cooke, T. E. (1989a) Disclosure in the corporate annual reports of Swedish companies, *Accounting and Business Research*, 19(74), pp. 113–124.

Cooke, T. E. (1989b) Voluntary corporate disclosure by Swedish companies, *Journal of International Financial Management and Accounting*, 1(2), pp. 171–195.

Cooke, T. E. (1991) An assessment of voluntary disclosure in the annual reports of Japanese corporations, *The International Journal of Accounting*, 26, pp. 174–189.

Cooke, T. E. (1992) The impact of size, stock market listing and industry type on disclosure in the annual reports of Japanese listed corporations, *Accounting and Business Research*, 22(87), pp. 229–237.

Cooke, T. E. (1993) Disclosure in Japanese corporate annual reports, *Journal of Business Finance and Accounting*, 20(4), pp. 521–535.

Cooke, T. E. (1998) Regression analysis in accounting disclosure studies, *Accounting and Business Research*, 28(3), pp. 209–224.

Copeland, R. M. and Fredericks, W. (1968) Extent of disclosure, *Journal of Accounting Research*, 6(1), pp. 106–113.

Core, J. E. (2001) A review of the empirical disclosure literature: discussion, *Journal of Accounting and Economics*, 31, pp. 441–456.

Cormier, D. *et al.* (1995) The auditor's consideration of the going concern assumption: a diagnostic model, *Journal of Accounting, Auditing and Finance*, 10(2), pp. 201–222.

Coy, D. *et al.* (1993) Quantifying the quality of tertiary education annual reports, *Accounting and Finance*, 33(2), pp. 121–130.

Depoers, F. (2000) A cost benefit study of voluntary disclosure: some empirical evidence from French listed companies, *European Accounting Review*, 9(2), pp. 245–263.

Downs, G. M. and Barnard, J. M. (2002) Clustering methods and their uses in computational chemistry, in: K. B. Lipkowitz and D. B. Boyd (Eds) *Reviews in Computational Chemistry*, Vol. 18, pp. 1–40 (New York: John Wiley).

Dumontier, P. and Raffournier, B. (1999) Vingt ans de recherche positive en comptabilité financière, *Comptabilité–Contrôle–Audit*, 5(May), pp. 179–197.

Emmanuel, C. and Garrod, N. (2004) Rules- versus judgement-based accounting disclosure in the UK, *Journal of Accounting and Public Policy*, 23(6), pp. 441–455.

Eng, L. L. and Mak, Y. T. (2003) Corporate governance and voluntary disclosure, *Journal of Accounting and Public Policy*, 22(4), pp. 325–345.

Entwistle, G. M. (1999) Exploring the R&D disclosure environment, *Accounting Horizons*, 13(4), pp. 323–341.

Ferguson, M. J. *et al.* (2002) Voluntary disclosure by state-owned enterprises listed on the stock exchange of Hong Kong, *Journal of International Financial Management and Accounting*, 13(2), pp. 125–152.

Firth, M. (1979) The impact of size, stock market listing, and auditors on voluntary disclosure in corporate annual reports, *Accounting and Business Research*, (Autumn), pp. 273–280.

Firth, M. (1984) The extent of voluntary disclosure in corporate annual reports and its association with security risk measures, *Applied Economics*, 16(2), pp. 269–277.

Giner, B. (1997) The influence of company characteristics and accounting regulation on information disclosed by Spanish firms, *The European Accounting Review*, 16(1), pp. 45–68.

Gray, R. *et al.* (2001) Social and environmental disclosure and corporate characteristics: a research-note and extension, *Journal of Business Finance and Accounting*, 28(3/4), pp. 327–356.

Haniffa, R. M. and Cooke, T. E. (2002) Culture, corporate governance and disclosure in Malaysian corporations, *Abacus*, 38(3), pp. 317–349.

Healy, P. M. and Palepu, K. G. (2001) Information asymmetry, corporate disclosure, and the capital markets: a review of the empirical disclosure literature, *Journal of Accounting and Economics*, 31, pp. 405–440.

Ho, S. and Wong, K. S. (2001) A study of the relationship between corporate governance structures and the extent of voluntary disclosure, *Journal of International Accounting, Auditing & Taxation*, 10(2), pp. 139–156.

Hossain, M. *et al.* (1994) Voluntary disclosure in an emerging capital market: some empirical evidence from companies listed on the Kuala Lumpur stock exchange, *The International Journal of Accounting*, 29(3), pp. 334–351.

Hossain, M. *et al.* (1995) Voluntary disclosure in the annual reports of New Zealand companies, *Journal of International Financial Management and Accounting*, 6(1), pp. 69–85.

IASC (1998) *International Accounting Standard (IAS) No. 37: Provisions, Contingent Liabilities and Contingent Assets* (London: International Accounting Standards Committee).

Iman, R. L. and Conover, W. J. (1979) The use of rank transformation in regression, *Technometrics*, 21, pp. 499–509.

Jaggi, B. and Low, P. Y. (2000) Impact of culture, market forces and legal system on financial disclosures, *The International Journal of Accounting*, 35(4), pp. 495–419.

Jensen, M. and Meckling, W. (1976) Theory of the firm: managerial behavior, agency costs and ownership structure, *Journal of Financial Economics*, 3(4), pp. 305–360.

Jones, J. (1991) Earnings management during import relief investigations, *Journal of Accounting Research*, 29(2), pp. 193–228.

Lang, M. H. and Lundholm, R. (1993) Cross-sectional determinants of analyst ratings of corporate disclosures, *Journal of Accounting Research*, 31(2), pp. 246–271.

Lang, M. H. and Lundholm, R. J. (1996) Corporate disclosure policy and analyst behavior, *The Accounting Review*, 71(4), pp. 467–492.

Leftwich, R. W. *et al.* (1981) Voluntary corporate disclosure: the case of interim reporting, *Journal of Accounting Research*, 19, pp. 50–77.

Leung, S. and Horwitz, B. N. (2004) Director ownership and voluntary segment disclosure: Hong Kong evidence, *Journal of International Financial Management and Accounting*, 15(3), pp. 235–260.

Malone, D. *et al.* (1993) An empirical investigation of the extent of corporate financial disclosurein the oil and gas industry, *Journal of Accounting, Auditing and Finance*, 3(3), pp. 249–273.

Marston, C. L. and Robson, P. (1997) Financial reporting in India: changes in disclosure over the period 1982 to 1990, *Asia-Pacific Journal of Accounting*, 4(1), pp. 109–139.

Marston, C. L. and Shrives, P. J. (1991) The use of disclosure indices in accounting research: a reviewarticle, *British Accounting Review*, 23(3), pp. 195–210.

McNally, G. M. *et al.* (1982) Corporate financial reporting in New Zealand: an analysis of user preferences, corporate characteristics and disclosure practices for discretionary information, *Accounting and Business Research*, 13(49), pp. 11–20.

Meek, G. K. *et al.* (1995) Factors influencing voluntary annual report disclosures by U.S., U.K. and continental European multinational corporations, *Journal of International Business Studies*, 26(3), pp. 555–572.

Milligan, G. W. and Cooper, M. C. (1986) A study of the comparability of external criteria for hierarchical cluster analysis, *Multivariate Behavioral Research*, 21(4), pp. 441–458.

Mitchell, J. D. *et al.* (1995) Voluntary disclosure of segment information: further Australian evidence, *Accounting and Finance*, 35(November), pp. 1–16.

Moore, M. L. and Buzby, S. (1972) The quality of corporate financial disclosure: a comment, *The Accounting Review*, 47(3), pp. 581–584.

Neter, J. et al. (1983) *Applied Linear Regression Models* (Homewood: Irwin).

Owusu-Ansah, S. (1998) The impact of corporate attributes on the extent of mandatory disclosure and reporting by listed companies in Zimbabwe, *The International Journal of Accounting*, 33(5), pp. 605–631.

Patton, J. and Zelenka, I. (1997) An empirical analysis of the determinants of the extent of disclosure in annual reports of joint stock companies in the Czech Republic, *European Accounting Review*, 6(4), pp. 606–626.

Prencipe, A. (2004) Proprietary costs and determinants of voluntary segment disclosure: evidence from Italian listed companies, *European Accounting Review*, 13(2), pp. 319–340.

Raffournier, B. (1995) The determinants of voluntary financial disclosure by Swiss listed companies, *European Accounting Review*, 4(2), pp. 261–280.

Salamon, G. and Dhaliwal, D. (1980) Company size and financial disclosure requirements with evidence from the segmental reporting issue, *Journal of Business Finance and Accounting*, 7(4), pp. 555–568.

Siegel, S. and Castellan, N. J. (1988) *Nonparametric Statistics for the Behavioral Sciences* (New York: McGraw-Hill).

Singhvi, S. (1968) Characteristics and implications of inadequate disclosure: a case study of India, *The International Journal of Accounting*, 3(2), pp. 29–43.

Singhvi, S. and Desai, H. (1971) An empirical analysis of the quality of corporate financial disclosure, *The Accounting Review*, 46(1), pp. 129–138.

Sneath, P. H. and Sokal, R. R. (1973) *Numerical Taxonomy* (San Francisco, CA: Freeman).

Stanga, K. G. (1976) Disclosure in published annual reports, *Financial Management*, (Winter), pp. 42–52.

Stolowy, H. and Tenenhaus, M. (1998) International accounting education in western Europe, *European Accounting Review*, 7(2), pp. 289–314.

Sucher, P. *et al.* (1999) The images of the big six audit firms in the Czech Republic, *European Accounting Review*, 8(3), pp. 499–521.

Tai, B. Y. K. *et al.* (1990) Non-compliance with disclosure requirements in financial statements: the case of Hong Kong companies, *The International Journal of Accounting*, 25(2), pp. 99–112.

Urban, D. L. (2004) Multivariate analysis – working with groups – how are the groups different?, March 22, 8 pp.

Wallace, R. S. O. and Naser, K. (1995) Firm-specific determinants of comprehensiveness of mandatory disclosure in the corporate annual reports of firms listed on the stock exchange of Hong Kong, *Journal of Accounting and Public Policy*, 14(4), pp. 311–368.

Wallace, R. S. O. *et al.* (1994) The relationship between the comprehensiveness of corporate annual reports and firm characteristics in Spain, *Accounting and Business Research*, 25(97), pp. 41–53.

Watson, A. *et al.* (2002) Voluntary disclosure of accounting ratios in the UK, *British Accounting Review*, 34(4), pp. 289–313.

Watts, R. and Zimmerman, J. (1978) Towards a positive theory of the determination of accounting standards, *The Accounting Review*, 53(1), pp. 112–134.

Williams, S. M. (1999) Voluntary environmental and social accounting disclosure practices in the Asia-Pacific region: an international empirical test of political economy theory, *The International Journal of Accounting*, 34(2), pp. 209–238.

X (1999) *Règles et méthodes relatives aux comptes consolidés, arrêté du 22 juin 1999 portant homologation du règlement 99-02 du 29 avril 1999 du comité de la réglementation comptable* [*Rules and Methods Relating to Consolidated Accounts – Decree of June 22, 1999, Approving the Regulation No. 99-02 of April 29, 1999, of the Accounting Regulation Committee*] (Paris: Accounting Regulation Committee).

Appendix A. Summary of disclosure studies

	Object of study	Country	Year	No. of firms (or obs.)	No. of disclosure items	Dependent variables	Main independent variables	Research design	Results
Singhvi (1968)	Extent of disclosure (generalist approach)	India	1963–65	45	34	Index (see Cerf), weighted items	Size, rate of return, earnings margin, audit firm, type of management, number of stockholders	Univariate	Size, management, number of stockholders
Singhvi and Desai (1971)	Extent of disclosure (generalist approach)	USA	1965	155	34	Index (see Cerf), weighted items	Size, number of shareholders, listing status, size of auditing firm, rate of return and earnings margin	Univariate Multivariate (linear regression)	Listing status
Buzby (1975)	Extent of disclosure (generalist approach)	USA	1970 or 1971	88	39	Index, weighted items	Size, listing status	Two matched samples Univariate	Size
Stanga (1976)	Extent of disclosure (generalist approach)	USA	1972 or 1973	80	79	Index, weighted items	Size, industry	Univariate Multivariate (linear regression)	Size, industry
Firth (1979)	Voluntary disclosure	UK	1976	180	48	Index, weighted items	Size, listing status, audit firm	Univariate	Size, listing status

(*continued*)

Appendix A. Continued

	Object of study	Country	Year	No. of firms (or obs.)	No. of disclosure items	Dependent variables	Main independent variables	Research design	Results
McNally *et al.* (1982)	Voluntary disclosure	New Zealand	1979	103	41	Index, weighted items	Financial characteristics (size, rate of return, growth), audit firm, industry	Univariate	Size
Firth (1984)	Voluntary disclosure	UK	1977	100	48	Disclosure index, weighted	Stock market risk	Linear regression	No significant relation
Chow and Wong-Boren (1987)	Extent of voluntary disclosure	Mexico	1982	52	24	Two scores: one weighted, one unweighted	Size, leverage, proportion of assets in place	Multivariate (linear regression)	Size
Cooke (1989a)	Extent of disclosure (mandatory and voluntary)	Sweden	1985	90	224	Disclosure index (unweighted)	Listing status, parent company relationship, size, number of shareholders	Multivariate Three regression models	Listing status, size
Cooke (1989b)	Extent of voluntary disclosure	Sweden	1985	90	146	Index (actual disclosure/ possible disclosure), unweighted items	Size, listing status, parent company relationship, industry	Univariate Multivariate (linear regression – stepwise)	Listing status
Tai *et al.* (1990)	Mandatory disclosure	Hong Kong	1987	76	11	Index additive (unweighted)	Size, industry, audit firm	Univariate	Size
Cooke (1991)	Voluntary disclosure	Japan	1988	48	106	Disclosure index (relative) (unweighted)	Size, listing status, industry	Univariate Multivariate Three regression models	Size

Cooke (1992)	Mandatory and voluntary disclosure	Japan	1988	35	165	Disclosure index (relative) (unweighted)	Size, listing status, industry	Multivariate (linear regression) Factor analysis of size variables	Size, listing status, industry
Cooke (1993)	Extent of voluntary disclosure	Japan	1988	48	195	Index, unweighted items	Listing status	Univariate	Listing status
Coy *et al.* (1993)	Tertiary education annual reports	New Zealand	1985–90	33	43	Two scores: unweighted and weighted ('Accountability Disclosure Score')	No variable	No analysis	–
Malone *et al.* (1993)	All financial disclosure in oil and gas industry	USA	1986	125	129	Weighted disclosure index	Size, listing status, leverage, profitability, audit firm	Stepwise regression model	Exchange listing status, ratio of debt to total equity, number of shareholders
Ahmed and Nicholls (1994)	Mandatory disclosure	Bangladesh	1988	63	94	Disclosure index (relative) (unweighted) (Cooke)	Size, leverage, audit firm, multinationality, qualification of the chief accountant	Univariate Multivariate (two regression models) (stepwise)	Multinationality, accountant's qualification, size
Hossain *et al.* (1994)	Voluntary disclosure	Malaysia	1991	67	78	Disclosure index (relative) (unweighted) (Cooke)	Size, ownership structure, leverage, assets-in-place, audit firm, listing status	Univariate Multivariate	Size, ownership structure, listing status
Wallace *et al.* (1994)	Mandatory and voluntary disclosure	Spain	1991	50	79	Disclosure index (unweighted)	Size, listing status, leverage, profitability, audit firm, liquidity	Multivariate (rank OLS regression)	Size (+), listing status (+); liquidity (−)

(*continued*)

Appendix A. Continued

	Object of study	Country	Year	No. of firms (or obs.)	No. of disclosure items	Dependent variables	Main independent variables	Research design	Results
Hossain *et al.* (1995)	Extent of voluntary disclosure	New Zealand	1991	55	95	Index (Cooke), unweighted items	Size, leverage, assets-in-place, audit firm, listing status	Multivariate (OLS)	Size, leverage, listing
Meek *et al.* (1995)	Voluntary disclosure	US, UK, France, Germany, Netherlands		116 + 64 + 16 + 12 + 18	85	Score (unweighted) Strategic, non-financial, financial	Size, country origin, industries, leverage, multinationality, profitability, listing status	Linear regression: four models (overall, strategic, non-financial and financial)	Size, country, listing status
Raffournier (1995)	Voluntary financial disclosure (generalist approach) Annual reports Listed firms	Switzerland	1991	161	30	Index (score, see Cooke, 1989a,b, 1992) (unweighted items)	Company size, leverage, profitability, ownership structure, internationality, auditor size, industry type	Univariate analyses and multiple linear regressions (stepwise)	Size, internationality
Wallace and Nasser (1995)	Extent of mandatory disclosure	Hong Kong	1991	80	142	Index, unweighted items	Foreign registered office, profit margin, earnings return, liquidity ratio, leverage, size, outside shareowners, conglomerates, audit firm	Multivariate: OLS and rank OLS regression	Size, conglomerates, profits
Ahmed (1996)	Mandatory and voluntary disclosure	Bangladesh	1987–88, 1992–93	118	150	Disclosure index, unweighted index	Size, leverage, audit firm, relation with parent, qualification of accountants	Regression: logarithm of the odds ratio	Audit, multinationality

Marston and Robson (1997)	Mandatory and voluntary disclosure	India	1983, 1990	29	17	Disclosure index (score = 1/0, or a scale based on judgment)	Size, change in time	Univariate	Size
Giner (1997)	Mandatory and voluntary disclosure	Spain	1989–91	138 obs.	50	Index (weighted items)	Size, listing status, profitability, leverage, audit firm, industry, dividend pay-out	Multivariate (stepwise regression)	Size, auditing, stock exchange, profitability
Patton and Zelenka (1997)	Extent of disclosure (generalist approach) Joint stock companies	Czech Republic	1993	50	37 + 12 + 17	Index (three levels of indexes) (unweighted items)	Size, performance, risk factors, other monitoring factors (listing status, big six auditing firms, industry)	Univariate analyses and multiple linear regressions No collinearity problem (VIF, condition indexes)	Type of auditor, number of employees
Owusu-Ansah (1998)	Mandatory disclosure	Zimbabwe	1994	49	214	Relative disclosure index (unweighted)	Size, ownership, age, multinational affiliation, profitability, audit, industry, liquidity	Multivariate: four regressions: OLS, rank OLS, without influential observations, robust	Size, ownership, age, multinational affiliation, profitability
Entwistle (1999)	R&D disclosure environment	Canada	1994 (or 1993 or 1995)	113	–	Content analysis (number of sentences)	R&D expense proportion, capitalization of R&D, cross-listing status, industry, capital structure, firm size	Multiple linear regression	R&D intensity, cross listing and industry

(*continued*)

Appendix A. Continued

	Object of study	Country	Year	No. of firms (or obs.)	No. of disclosure items	Dependent variables	Main independent variables	Research design	Results
Williams (1999)	Voluntary environmental and social disclosure	Seven Asia-Pacific nations	1995	356	–	Content analysis (number of sentences)	Culture, political and civil system, legal system, level of economic development, equity market, control variables	Three linear regressions	Uncertainty avoidance, masculinity, political and civil systems
Chen and Jaggi (2000)	Mandatory disclosure	Hong Kong	1993, 1994	87	142	Disclosure index (unweighted) (see Wallace and Naser, 1995)	Independent non-executive directors, family control, profitability, leverage, size, audit firm	OLS regression	Independent non-executive directors
Depoers (2000)	Voluntary disclosure (generalist approach)	France	1995	102	65	Disclosure score (unweighted)	Firm size, foreign activity, ownership structure, leverage, size of auditing, proprietary costs related to competition, labor pressure	Multiple linear regression (two OLS regressions to avoid collinearity problems with the high correlation between size and barriers to entry) Stepwise procedure	Foreign activity and size
Jaggi and Low (2000)	Mandatory and voluntary disclosure	28 countries	1991	28	90	Relative disclosure index (unweighted)	Cultural, legal and financial variables	Univariate, multivariate (six regression models)	Common law, culture

Gray *et al.* (2001)	Social and environmental disclosure	UK	1998–95	100	–	Eight measures of disclosure (CSEAR Social and Environmental Disclosure Database)	Profit, turnover, capital employed, industry classification, number of employees	Eight OLS regressions	No unique and stable relationship
Ho and Wong (2001)	Voluntary disclosure	Hong Kong	1998	98	20	Relative disclosure index (weighted items)	Independent non-executive directors, audit committee, dominant personalities, family + control variables	Multivariate (linear regression)	Audit committee, family
Bujaki and McConomy (2002)	Voluntary disclosure	Canada	1997	272	25	Disclosure index	Financial condition, leverage, share issue, unrelated directors, regulated industries, medium, size	Linear regression	Unrelated directors, leverage
Chau and Gray (2002)	Voluntary disclosure	Hong Kong Singapore	1997	62	Approx. 110	Disclosure index (unweighted) (three scores)	Ownership structure, size, leverage, audit firm, profitability, multinationality, industry	Multivariate (linear regression)	Ownership structure
Ferguson *et al.* (2002)	Voluntary disclosure	Hong Kong	1995–96	142	93	Disclosure index Gray *et al.* (2001), Meek *et al.* (1995) Unweighted	Firm type, size, leverage, industry, listing status	Univariate Multivariate (linear regression) Total score. Replication with partition: strategic, non-financial, financial information	Firm type, leverage (type of disclosure)

(*continued*)

Appendix A. Continued

	Object of study	Country	Year	No. of firms (or obs.)	No. of disclosure items	Dependent variables	Main independent variables	Research design	Results
Haniffa and Cooke (2002)	Voluntary disclosure	Malaysia	1995	167	65	Disclosure index (unweighted)	Corporate governance, cultural and firm-specific	Linear regression (31 variables) Restricted model	Family members sitting on board, non-executive chairman
Archambault and Archambault (2003)	Voluntary and non-voluntary disclosure	33 countries	1992, 1993	621	85	Disclosure index (unweighted)	Culture, national, financial systems	Multivariate (linear regression)	Many factors
Eng and Mak (2003)	Voluntary disclosure	Singapore	1995	158	84	Aggregated disclosure score of non-mandatory strategic, non-financial and financial information	Ownership structure, board composition	OLS regression (one model + three variations)	Lower managerial ownership, government ownership, outside directors, lower debt
Prencipe (2004)	Extent of voluntary segment disclosure	Italy	1997	64	9	Disclosure index (unweighted and weighted)	Correspondence between the segments and legally identifiable sub-groups of companies, growth rate, listing status, age, size, ownership diffusion, financial leverage, profitability	OLS regression (two models)	Correspondence and listing status age
Cahan *et al.* (2005)	Voluntary disclosure	17 countries	1998 or 1999	216	Botosan's (1997) index	Disclosure index	Global diversification	OLS regression	Global diversification, number of analysts, size

Appendix B. List of sample firms showing disclosure indexes for 2001

Company	Index	Company	Index
Accor	0.643	L'Oréal	0.500
Air France	0.786	LVMH	0.214
Air Liquide	0.214	M6	0.500
Alcatel	0.643	Marionnaud	0.571
Alstom	0.571	Michelin	0.571
Alten	0.500	Neopost	0.643
Altran	0.143	Nexans	0.571
Atos Origin	0.786	NRJ	0.214
Aventis	0.929	Oberthur	0.429
Béghin-Say	0.500	Orange	0.143
Bic	0.357	Pechiney	0.643
Bouygues	0.857	Penauille	0.786
Business Objects	0.429	Pernod Ricard	0.571
Cap Gemini	0.214	PPR	0.714
Carbone Lorraine	0.500	Provimi	0.357
Carrefour	0.286	PSA	0.714
Casino	0.571	Publicis	0.714
Castorama	0.214	Renault	0.571
Céréol	0.357	Rexel	0.643
CGG	0.357	Rhodia	0.500
Ciments français	0.500	Royal Canin	0.571
Clarins	0.429	Sagem	0.286
Club Med	0.643	Saint-Gobain	0.429
Cointreau	0.500	Sanofi	0.571
Danone	0.429	Schneider	0.571
Dassault Systèmes	0.500	Seb	0.643
EADS	0.500	Simco	0.571
Eiffage	0.357	Sodexho	0.429
Elior	0.643	Soitec	0.286
Equant	0.286	Sophia	0.286
Essilor	0.429	Sopra Group	0.429
Eurodisney	0.143	Spir	0.571
Eurotunnel	0.286	SR Telecom	0.500
Faurecia	0.714	Steria Groupe	0.357
France Telecom	0.571	Technip	0.643
Galeries Lafayette	0.357	TF1	0.643
Gecina	0.643	Thales	0.643
Gemplus	0.357	Thomson	0.786
GFI	0.500	Total	0.786
GrandVision	0.357	Ubisoft	0.429
Guyenne et Gascogne	0.429	Unilog	0.714
Haulotte	0.500	Valeo	0.714
Havas	0.429	Vallourec	0.786
Hermes	0.429	Vinci	0.714

(*continued*)

Appendix B. Continued

Company	Index	Company	Index
Imerys	0.857	Vivendi Environnement	0.500
Infogrames	0.286	Vivendi Universal	0.429
Ingenico	0.286	Wanadoo	0.357
JCDecaux	0.500	Wavecom	0.643
Lafarge	0.643	Wendel	0.500
Lagardère	0.286	Zodiac	0.571

Disclosure based on 14 items.

Appendix C. List of the 14 surveyed items

1. Inclusion in the 'accounting principles' part of the notes to the financial statements (the 'notes' in the rest of this paper) of a specific note on the treatment of provisions (excluding pension liabilities).
2. Inclusion in the notes of the description of a particular type of provision (e.g. section on the provisions for restructuring, for repair work, etc.).
3. Inclusion in the notes of a description of the methods for computation of provisions (e.g. statistical determination of a given provision).
4. Inclusion in the 'accounting principles' note of a note on the treatment of pension liabilities.
5. Reference in the notes concerning pension liabilities to the accounting standard applied (IAS 19, FAS 87, French regulation 99-02).
6. Inclusion in the notes concerning pension liabilities of details of the computation of provisions. (One point is attributed when the notes disclose information such as statistical data on the population, discount rate, amount of funded pension liability, etc.)
7. Disclosure in the 'Equity and liabilities' side of the balance sheet of a specific line for 'Provisions'. (We attribute a 0 when the company reports a heading such as 'Other long-term debts' covering provisions and other debts.)
8. Disclosure in the balance sheet of a Year X1/Year X0 comparison.[a]
9. Disclosure in the balance sheet of a Year X1/Year X0/Year X − 1 comparison. (The disclosure of the two preceding years' figures will become compulsory in 2005 with implementation of the IFRS. All companies which have already incorporated this change were attributed one point.)
10. Inclusion in the notes of a statement of changes in provisions (increases, decreases, etc.).
11. Breakdown given of the item 'Other' (provisions) allowing for restatement and distribution of this item between the other major provisions. One point is attributed when the company discloses quantified information.
12. Distinction between short-term and long-term provisions in the notes.
13. Disclosure of specific lines of provisions. If there is only a breakdown between 'risks' and 'expenses', a 0 is given. If more lines (with a threshold of five) are disclosed, a 1 is given.

(*continued*)

Appendix C. Continued

14.	Amount relating to 'other provisions' lower than the average for the 100 companies. The importance of the 'other' line is a major factor in the transparency of financial information disclosed by the company. The higher it is (measured relative to the size = assets of the firm), the greater the possibilities of accounts manipulation are. We have transformed this complex variable into a dummy variable, specifying that all companies with an 'other' lower than the sample average would be attributed one point, and the rest 0.

[a]This item is dropped in the DIV analysis, because it is the only one which can be considered as mandatory and is published by all firms in our sample.

Appendix D. DIV Analysis Method Explained

In order to separate objects with DIV, a within-cluster variance criterion is defined. Let N be the number of objects in set Ω. All objects are described on p real value variables by vector $x_i \in R^p$, $i = 1, \ldots, N$. Each object is weighted by a real value $w_i \geq 0$ $(i = 1, \ldots, N)$. The weights are usually equal to 1 in classic data analysis.

A within-cluster variance of a cluster C_k is given by

$$I(C_k) := \sum_{x_i \in C_k} w_i \left\| x_i - \bar{x}_{C_k} \right\|^2$$

where $\bar{x}_{C_k}$ is the centroid of the cluster C_k

$$\bar{x}_{C_k} = \mu_k \sum_{x_i \in C_k} w_i x_i$$

$$\mu_k = \sum_{x_i \in C_k} w_i.$$

The variance criterion of K-partition $C = (C_1, \mathrm{K}, C_K)$ is given by

$$W(C) := \sum_{i=1}^{K} I(C_i).$$

This criterion is minimized among bipartitions induced by a set of binary questions. Let C_i be a set of n_i objects. The goal is to find the bipartition $C_i = (C_i^1, C_i^2)$ such that the within-cluster inertia is minimum. In DIV, a cluster C is divided according to a binary question of the form 'Is $Y_j \leq c$?' where Y_j is a real variable and $c \in R$ is called the cut point. The bipartition $C_i = (C_i^1, C_i^2)$ induced by the binary question is defined as follows. Let x be an object in C_i. If $Y_j(x) \leq c$ then $x \in C_i^1$ else $x \in C_i^2$. Those objects in C answering 'yes' go to the left descendant cluster and those answering 'no' to the right descendant cluster.

Let $P_K = (C_1, \mathrm{K}, C_K)$ be a K-partition of Ω. At each stage, a new $(K+1)$-partition is obtained by separating one of the clusters $C_i \in P_K$ in two new clusters C_i^1 and C_i^2. The new partition is chosen such that $P_{K+1} = P_K \cup \{C_i^1, C_i^2\} - \{C_i\}$ has minimum within-cluster variance. Since $W(P_{K+1}) = W(P_K) - I(C_i) + I(C_i^1) + I(C_i^2)$, minimizing $W(P_{K+1})$ is equivalent to choosing the cluster so that the difference between the within-cluster variance of C_i and the within-cluster variance of its bipartition (C_i^1, C_i^2) is maximum. Thus, the criterion for selecting the cluster to split is given by

$$\Delta(C_i) = I(C_i) - I(C_i^1) - I(C_i^2).$$

This means that the bipartitions of all the clusters of the partition P_K have been defined previously. At each stage, the bipartitions of the two new clusters (C_i^1, C_i^2) are defined and used in the next stage.

The divisions are stopped after a number L of iterations and L is given as input by the user. The last partition obtained in the last iteration is an $L+1$-clusters-partition. The stopping rule ensures that the partitions of smallest within-cluster variance of the total hierarchy are obtained after L iterations. However, this stopping rule does not solve the issue of determining the number of clusters in the data-set (Milligan and Cooper, 1986).

The output of this divisive clustering method is a hierarchy H whose singletons are the $L+1$ clusters of the partition obtained in the last iteration of the algorithm. Each cluster $C_k \in H$ is indexed by $\Delta(C_k)$. Because Δ is a non-decreasing mapping, $C_k \subset C_{k'} \Rightarrow \Delta(C_k) \leq \Delta(C_{k'})$, there will be no inversions in the dendrogram of the hierarchy. This hierarchy is also a decision tree. The L clusters are the leaves and the nodes are the binary questions selected by the algorithm. Each cluster is characterized by a rule defined according to the binary questions leading from the root to the corresponding leaves.