A STUDY OF THE ABILITY OF (PARTIALLY) AUTOMATED DISCLOSURE SCORES TO EXPLAIN THE INFORMATION CONTENT OF ANNUAL REPORT NARRATIVES FOR FUTURE EARNINGS

A thesis submitted to the University of Manchester for the degree of
Doctor of Philosophy in the Faculty of Social Science and Law

2004

KHALED SAID MOHAMED HUSSAINEY

MANCHESTER SCHOOL OF ACCOUNTING AND FINANCE

Table of Contents

List of Tables

List of Figures

List of Abbreviations

%	Percentage
£	Sterling pound
£M	Millions of Sterling pounds
00E	Year 2000 estimated
00F	Year 2000 forecasted
ACCA	Association of Chartered Accountants
ADR	American Depository Receipts
AIMR-FAF	Association for Investment Management Research - Financial analysts Federation
ASB	Accounting Standards Board
ASSC	Accounting Standards Steering Committee
CA	Canada
CAPM	Capital Asset Pricing Model
CEO	Chief Executive Officer
CEOs	Chief Executive Officers
CFPS	Cash Flow Per Share
CFROI	Cash Flow Return on Investment
CIFAR	Center for International Financial Analysis and Research
COGS	Cost of Goods Sold
CROCI	Cash Return On Capital Invested
D	Dummy variable
D/E	Dividend/Earnings
DCF	Dividend Cash Flow
DM	Deutsche Mark
DPS	Dividend Per Share

DOPF	Operating and Financial Review Projections Index
DOFR	Operating and Financial Review Index
DSEG	Segmental Reporting Index
e.g.	Exempli gratia (Latin: for instance)
EBDIT	Earning Before Debt Interest and Tax
EBDITA	Earning Before Debt Interest, Tax and Amortisation
EBIT	Earning Before Interest and Tax
EBITA	Earning Before Interest, Tax and Amortisation
EBITDA	Earnings Before Interest, Tax, Depreciation and Amortisation
EBT	Earnings Before Tax
E-commerce	Electronic Commerce
Eds.	Editors
EP ADR	Earnings/Price American Depositary Receipts
EPS	Earnings Per Share
ERC	Earnings Response Coefficient
et al.	et alia (Latin: and other)
EU	Europe
EV	Enterprise Value
EVA	Economic Value Added
FCF	Free Cash Flow
FERF	The Financial Executive Research Foundation
FRS	Financial Reporting Standard
FTSE	Financial Times Stock Exchange
FY	Financial Year
GAAP	General Accepted Accounting Principles
GDP	Gross Domestic Product

H1	Half Year
H1:	Hypothesis 1
H2	Hypothesis 2
i	Company i
i.e.	id est (Latin: that is to say)
ICAS	The Institute of Chartered Accountants of Scotland
ICAEW	The Institute of Chartered Accountants of England and Wales
IR	Investor relations
IRR	Internal Rate of Return
IT	Information Technology
MBAR	Market Based Accounting Research
MD&A	Management's Discussion and Analysis
MPhil	Master of Philosophy
MSc	Master of Science
MV	Market Value
MVA	Market Value Added
N	Number of Observations
N5	*Nudist* Version 5
NA	Not Applicable
No.	Number
NOPAT	Net Operating Profit After Tax
NPAT	Net Profit After Tax
NPV	Net Present Value
NUDIST	Non-numerical Unstructured Data Indexing Searching & Theorizing
NWC	Net Working Capital
OECD	Organisation for Economic Co-operation and Development

OFR	Operating and Financing Review
OpFCF	Operating Free Cash Flow
P	Price
p.	Page
PBT	Profit Before Tax
PE	Price/Earnings
PhD	Philosophiae Doctor (Latin: doctor of philosophy)
PLC	Public Limited Company
PTP	Pre Tax Profit
PV	Present Value
QSR	Qualitative Solutions and Research
R&D	Research and Development
ROA	Return On Assets
ROACE	Return On Average Capital Employed
ROCE	Return On Capital Employed
ROE	Return On Equity
ROI	Return On Investment
ROIC	Return On Invested Capital
SAS	Statistical Analytical Software
SEC	The Securities and Exchange Commission
SMAC	Society of Management Accountants of Canada
SPSS	Superior Performing Software Systems
t	Period t or year-end of period t
UK	United Kingdom
UMIST	University of Manchester Institute of Science and Technology
US	United States

USA	United States of America
VAT	Value Added Tax
WACC	Weighted Average Cost of Capital
WSJ	Wall Street Journal

Abstract

The study presents a new methodology for evaluating corporate voluntary disclosures in annual report narratives. Based on a new dataset of electronic annual reports and a standard text analysis software package, I text-search a large number of annual reports at minimal (marginal) cost. The resulting sample sizes are comparable to those employed in studies based on the AIMR-FAF database. A major advantage of this scoring methodology is that it is adaptable to the particular requirements of the research design. The importance of this feature is demonstrated when examining the association between prices leading earnings and the quality of corporate disclosure. To measure this association, I use the regression model of Collins et al. (1994).

While the study is unable to find any significant association between prices leading earnings and disclosure quality based on an all-inclusive topics list, this result reverses once I focus on a more narrowly defined metric based on forward-looking profit statements. The study also finds that changes in disclosure quality are positively related to changes in the importance of future earnings news for current returns. This effect is particularly strong when forward-looking statements are directly related to profit topics and when the time horizon in each statement is clearly specified.

Declaration

I declare that no portion of this thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Khaled Hussainey
10 June 2004

Copyright and Ownership of Intellectual Property Rights

Acknowledgements

I would like to thank the Almighty *God*, the most *Beneficent* and the most *Merciful*, who has empowered and enabled me to complete this thesis successfully.

I wish to give special thanks and to express on my sincerest gratitude to my supervisors, Professor Martin Walker and Dr. Thomas Schleicher, for their guidance, patience, time and enthusiasm throughout the stages of my PhD. Without their continuous assistance and encouragement, this thesis could not have been completed.

I would like to thank the editor of Accounting and Business Research (Professor Ken Peasnell) and an anonymous referee for their helpful comments on my paper.

Very special thanks go to my internal examiner, Professor Stuart Turley, and my external examiner, Professor Vivien Beattie (Glasgow University), for the time they took to read and comment on my thesis, as well as for their encouragement and positive input.

I am delighted to give my special appreciation to all administrative and academic staff in Manchester School of Accounting and Finance for their support. I take this opportunity to present a special appreciation to Dr. Mahbub Zaman (PhD Program Director) and Nicola Lord (PhD Administrator).

My colleagues of the School of Accounting and Finance of the University of Manchester were responsible for creating a pleasant working environment during the four years I was allowed to spend with them. I would like to express my special thanks to Abdalla Saleh, Fadi Alkaraan, Haitham Nobanee, Mohamed Sherif, Ali Nodel, Ali Elharidy and Sulaiman Mouselli. They played a major role in my development during these years, both on a scientific and a personal level.

I gratefully acknowledge the financial support given to me from the Egyptian Ministry of Education. Special thanks also should go to all administrative and academic staff in Department of Accounting & Auditing, Faculty of Commerce, Ain Shams University for their moral support.

I would like to thank my parents, my sisters and my brothers for their continuous moral support and encouragement during my study. Without that, the completion of this thesis would have probably taken much longer.

Last, but by no means least, I would like to thank my wife, Nahla, for her unlimited support, patience, sacrifice and enthusiasm during the course of my degree research. I am very grateful to her. My children, Yasser and Osama, have kept me cheerful all the time by their playfulness, smile and laughter. I love them dearly and I hope they will be successful in their life.

The Author

Khaled S. M. Hussainey was born in Egypt on Novermber, 1, 1972. He attended Ain Shams University in Egypt, graduating in June 1994 with a Bachelor of Commerce in Accounting with honour degree. Mr. Hussainey was awarded a scholarship from the Egyptian Ministry of Education to study M.Sc. and Ph.D. in Accounting and Finance. In September 1999, he pursued Accounting and Finance as his major field of study at the University of Birmingham in the UK, receiving a Master of Science in Accounting and Finance in September 2000. In September 2000, he registered for the Ph.D. degree in Manchester School of Accounting and Finance under the supervision of Professor Martin Walker and Dr. Thomas Schleicher. During his PhD study, Mr. Hussainey published a paper, with Thomas Schleicher and Martin Walker, in *Accounting and Business Research* in 2003.

From 1995 through 2001, Mr. Hussainey joined the Department of Accounting and Auditing, Faculty of Commerce, Ain Shams University as a tutor. From 2002 till now, he joined the Department of Accounting and Auditing at the same University as an Assistant Lecturer of Accounting.

Chapter 1: Introduction

1.1. Overview

Reported earnings have attracted major interest in the Market Based Accounting Research (MBAR) literature since the publication of two remarkable papers in the mid-1960s by Ball and Brown (1968) and Beaver (1968). The MBAR literature emerges as a response by accounting academics to the success of finance research on market efficiency. Market efficiency is concerned with the extent to which security prices fully reflect all available information. Since accounting is an important source of value-relevant information about companies, it was natural for researchers to study the efficiency of the stock market with respect to accounting information.

For the purposes of the present study, an important landmark article in the MBAR literature is Lev (1989). Lev (1989) surveys published papers of the relation between stock returns and earnings changes. He notes that the R^2 values obtained by regressing annual stock returns on annual earnings or earnings changes are very low (10% or lower). He also notes that the earnings response coefficient (ERC) values are very low (around 2.0 or lower). Lev ascribes these weak results to the low quality of accounting earnings.

The findings in Lev's survey article challenge researchers to identify the potential explanations for the weak contemporaneous return-earnings association. There are four main potential reasons for this weak association. These are: (1) market (informational) inefficiency, (2) the stock market reacts to value-relevant information that is not observed by the company (see Dye and Sridhar, 2002), (3) noise in reporting earnings and (4) earnings' lack of timeliness.

The present study focuses only on *one* of these potential reasons, the lack of timeliness in earnings. The literature on earnings' timeliness is concerned with the extent to which the market has access to value-relevant information other than reported earnings. Reported earnings lack timeliness due to the fact that many value-relevant events are reflected in stock prices as soon as the information reaches the market while their influence on reported earnings often occurs with a time lag. This lagged response of earnings is partly due to certain accounting conventions such as objectivity, conservatism and verifiability that prevent earnings from reflecting the value-relevant information on a timely basis.

Collins et al. (1994) is a major attempt to empirically investigate the reasons for the weak contemporaneous return-earnings association. They show that earnings' lack of timeliness is the most important contributing factor to the low contemporaneous return-earnings relation. The development by Collins et al. (1994) of a reliable measure of earnings timeliness makes it possible to measure the consequences of differential levels of corporate disclosures.

This chapter is structured as follows. Section 1.2 outlines the nature of the study. Section 1.3 addresses the motivations for carrying out this study. Research objectives and the main contributions of this study are presented in Section 1.4. Section 1.5 presents the outline of the thesis.

1.2. Nature of the Study

The present study contributes to the literature that relates to the economic consequences of corporate voluntary disclosures. The study is partially a response to the argument of Core (2001), who recommends that researchers can import techniques in natural

language processing from fields like computer science to construct their disclosure scores.

In the UK, the availability of an electronic data has improved recently. Consequently, it is possible to use computer software packages to score corporate disclosures. The study uses *Nudist* (Non-numerical Unstructured Data Indexing Searching and Theorizing) to automatically generate disclosure scores for large numbers of annual reports at minimal time, cost and effort.

Generally, producing automated disclosure scores for large samples of annual reports proceeds in three stages. The first is the construction of the list of disclosure items. Disclosure topics are constructed based on examining the textual content of analysts' reports. In this stage, a mixture of computer-aided content analysis and labour-intensive content analysis is involved. The computer analysis is used principally to identify and calculate the frequency of forward-looking sentences associated with every searched key word. Labour-intensive content analysis is used to identify the key noun in each forward-looking sentence. In the second stage, I identify key words that are frequently associated with forward-looking information in annual report narratives. In the final stage, the number of sentences for each firm that are forward-looking in nature and include a relevant topic is automatically counted via *Nudist*. The automated disclosure scores are then used examine the economic consequences of corporate voluntary disclosures in the UK.

Most of the published papers on the economic consequences of corporate disclosures are based on US data. In part, this is because the fact that it is easy to access AIMR-FAF (Association for Investment Management Research – Financial Analysts Federation) analyst disclosure ratings. In the UK, there are no analyst ratings like

AIMR-FAF. As a result, UK researchers use the labour-intensive content analysis approach to select the list of disclosure items and to score annual reports. This approach is extremely time consuming. This is specifically the case when analysing large samples of annual reports. As a result, to date there is no an empirical large-scale disclosure study undertaken in the UK.

The automated disclosure methodology developed in the present study allows me to participate actively in the empirical disclosure studies. In particular, it enables me to examine the association between the quality of forward-looking information in annual report narratives and prices leading earnings for a large sample of UK firms.

The association between corporate voluntary disclosures and 'prices leading earnings' has recently attracted major interest in US accounting journals. Following Lundholm and Myers (2002) and Gelb and Zarowin (2002), the present study investigates whether higher levels of forward-looking information in annual report narratives lead to a significant improvement in investors' forecasts of future earnings growth. In particular, if a firm reveals forward-looking information in its annual report, then this information should be reflected in stock prices. As a result, one would expect that realised future earnings will be partially anticipated by investors.

The present study seeks to test the hypothesis that high levels of forward-looking information (based on an all-inclusive topic list) in annual report narratives are associated with stock prices that are more informative about future earnings growth. While an all-inclusive forward-looking topics list might be a fair reflection of the forward-looking discussion in analysts' reports, it is clear that many included topics are weakly correlated with future earnings growth. Therefore, the study seeks to test the second hypothesis that high levels of forward-looking *Profit* information in annual

report narratives are associated with stock prices that are more informative about future earnings growth. The cross-sectional findings suggest that the ability of the market to anticipate future earnings changes is positively related to the quality of disclosure in the firm's annual report narratives. This effect is clearly significant when disclosure quality is defined in terms of forward-looking *profit* statements.

The study also examines whether ***changes*** in corporate disclosure change the extent to which stock returns anticipate future earnings changes. The results show that the disclosure practices of firms tend to be very persistent. As a result, it is difficult to identify the effects of changes in disclosure on changes in earnings' timeliness by an automatically generated disclosure measure. However, I find a positive effect of changes in disclosure on changes in earnings' timeliness only when I manually identify forward-looking sentences that are directly related to future profits and are referred to a well-specified period of time.

1.3. Motivations

The first motivation for this study is to address some of the gaps left by the disclosure indices literature. Prior research criticises the use of qualitative research methods, such as questionnaires and interviews, to obtain analyst's views regarding the importance of disclosure topics in the annual report (see, for example, Rogers and Grant, 1997). In addition, prior research also criticises the use of the annual report in the construction of disclosure indices (Schadewitz and Kanto, 1997). Consequently, I produce a new methodology for constructing disclosure indices. This methodology is based on what analysts actually include in their reports rather than on their opinions regarding the importance of particular annual report disclosure items. The study uses a structured

content analysis of the textual content of analysts' reports to produce a list of disclosure items.

The second motivation for the study arises from the unavailability of subjective disclosure ratings in the UK similar to AIMR-FAF ratings. UK researchers use labour-intensive content analysis methods to construct their disclosure indices. These involve a considerable amount of time and effort on the part of the researchers. Accordingly, their sample sizes are often small in comparison with US studies. Therefore, I aim to produce a new methodology for automating the generation of disclosure scores for large samples of annual reports with minimum cost and effort.

In line with the first and second motivations, the third motivation for the study arises from the inadequate use of the computer in undertaking accounting-based qualitative research (Core, 2001). The present study uses *Nudist* in evaluating the quality of corporate disclosure. *Nudist* offers advantages relative to labour-intensive scoring methods. These advantages include: (1) the ability to score very large numbers of annual reports at very low marginal costs, (2) an increase in the comparability of the disclosure scores across firms and over time, and (3) the ability to replicate the scores easily in subsequent disclosure studies.

Numerous studies examine the economic benefits of corporate voluntary disclosures in a variety of contexts. These contexts include prices leading earnings, cost of capital and analyst behaviour.[1] All these studies are conducted in US data where AIMR-FAF disclosure ratings are available. However, no UK study links the quality of corporate disclosure with any one of these contexts for large samples of annual reports. Therefore,

[1] Analyst behaviour refers to two major issues, which are analyst following and the characteristics of analysts' earnings forecasts. The characteristics of analyst forecasts are measured by forecast accuracy, the degree of dispersion among forecasts and the variability of forecast revisions during the year.

the fourth motivation is to undertake a large-scale disclosure study based on UK data. It is of interest to know whether the findings of US studies are replicated in the UK, an equally well-developed industrial country with different requirements with respect to financial reporting regulations.

The final motivation arises from the inability of previous studies to explain why there is no apparent association between annual report disclosures and 'prices leading earnings' (e.g., Gelb and Zarowin, 2002). The present study provides a partial answer to this question. It also highlights the importance of individual classes of information that contribute most to the phenomenon of prices leading earnings.

1.4. Research Objectives and Contributions

The research objectives of the study are: (1) to develop a new methodology for constructing disclosure indices, (2) to develop a new methodology for automating the generation of disclosure scores for large numbers of firms, and (3) to empirically examine the association between the quality of annual report narratives and prices leading earnings for large numbers of UK non-financial firms

There are a number of novel contributions this study makes to the literature. The first substantive contribution of the study is the use of a new method for constructing disclosure indices. This method is based on the analysis of the content of 60 analysts' reports from different brokerage houses. The resulting scoring sheet is presented in Chapter 4. This index contains 500 topics that represent the topics of interest to financial analysts in their decision-making process.

In the UK, there are no analyst ratings similar to those in the US. In addition, AIMR-FAF ratings were discontinued in 1997 (the last year of the disclosure scores was 1995). Therefore, the second major contribution of this study is a methodology for automating

the generation of disclosure scores for large samples of annual reports. Other researchers can easily replicate the resulting disclosure scores in subsequent disclosure studies. Chapter 5 presents the necessary stages which one needs to replicate this methodology in the future.

Based on the automated disclosure scores generated in Chapter 5, the present study is the first to undertake a large-scale disclosure study based on UK company data.[2] The study examines the association between the quality of forward-looking information in annual report narratives and prices leading earnings. The empirical results provide evidence that forward-looking profit information helps the market to anticipate future earnings changes more accurately. Identifying individual classes of information that contribute most to the phenomenon of prices leading earnings is the third major contribution of this study. Chapters 6 and 7 present the empirical findings.

1.5. Outline of the Study

Chapter 2 discusses the importance of annual report narratives. It also discusses the definition of forward-looking information and the advantages and the disadvantages of publishing forward-looking information in the annual report. In addition, the chapter reviews the proxies for measuring the quality of corporate disclosures. The chapter ends by discussing the difference between the traditional and the computerised content analysis approaches.

Chapter 3 reviews the academic literature concerning two research fields: (1) return-earnings regression models and (2) the economic consequences of corporate disclosures. The chapter starts by outlining the main published papers regarding prices leading earnings. It also provides more discussion about the theoretical model of Collins et al.

[2] The first results of this work are published in Hussainey et al. (2003). This thesis presents these results along with further results.

(1994), which is used in performing the regression analyses. In addition, the chapter reviews the empirical studies that investigate the effect of corporate disclosures on analyst behaviour, cost of equity and prices leading earnings. The chapter ends by introducing the extent to which the present study fills the gaps left by prior studies.

Chapter 4 describes the methodology that is adopted in the present study to create the list of disclosure items. This methodology is based on text-searching analysts' reports for the key words that appear relevant for financial analysts' decision-making processes. Based on a list of forward-looking key words, *Nudist* software is used to extract the forward-looking sentences that are published in analysts' reports. Each sentence is carefully read to identify the key topic(s). The scoring sheet represents the key topics appearing in the forward-looking sentences that are published in analysts' reports.

Chapter 5 extends the work of Chapter 4 by using the disclosure topics to score large samples of annual reports. One important feature of the scoring methodology used in the chapter is the automation of disclosure scores for large samples of annual reports through the 'Text Search' function in *Nudist*. The chapter ends by testing the reliability and the validity of my disclosure index.

Chapter 6 uses the cross-sectional regression analysis to examine the relation between the level of corporate disclosure and prices leading earnings. It discusses the research hypotheses, the empirical model, the data, and the variables definition. It also presents the descriptive statistics for the variables used in the analysis. The chapter ends by discussing the interpretation of the empirical results. In the chapter I find a significant evidence of share price anticipation of earnings in my sample of UK firms. My test for the association between my most general measure of corporate disclosure and share price anticipation of earnings yields insignificant results. However, significant results

are obtained with a more narrowly defined measure of disclosure that focuses specifically on forward-looking profit statements.

Chapter 7 examines the association between *changes* in disclosure quality and changes in earnings' timeliness. It includes the development of the research hypothesis and the sample selection. The results in the chapter suggest that my computer-based scoring methodology is not yet powerful enough to distinguish the changes in disclosure that are value-relevant, from those that are not. For this analysis I find that a less computer-intensive approach is more successful.

Chapter 8 summarises the main findings, the implication of the results, the limitations of the study and the suggestions for future research.

Chapter 2: Narrative Disclosures in Corporate Annual Reports: A Review

2.1. Overview

Demand for corporate disclosure arises from information asymmetry and agency conflicts between corporate managers, outside investors and intermediaries (Kothari and Short, 2003). Increasing the level of corporate disclosure reduces information asymmetry (Glosten and Milgrom, 1985). A rich information environment and low information asymmetry have many desirable consequences. One of these consequences is the increase in the market's ability to anticipate future earnings changes.

The present study reports the findings of investigating the association between the quality of annual report narratives and prices leading earnings for UK companies. The study uses a structural content analysis of analysts' reports to produce a list of disclosure items. It uses *Nudist* to assist in the identification of the disclosure items. Finally, annual report narratives are automatically scored by using the command files in *Nudist*.

The purpose of this chapter is to review the research on the quality of corporate voluntary disclosures in the annual report discussion section. Section 2.2 starts with presenting the importance of corporate annual reports to financial analysts. Specifically, this review discusses the importance of corporate narrative reporting. Section 2.3 emphasises the usefulness of forward-looking information published in the corporate annual reports. It also includes the definition of forward-looking information and the arguments for and against this type of information. Section 2.4 summarises the methodologies used to measure the quality of corporate disclosures. This summary reviews three major proxies for disclosure quality. These are: management forecasts, self-constructed disclosure indices, and subjective ratings such as 'AIMR-FAF disclosure ratings'. Section 2.5 discusses issues relating to the use of the content

analysis methodology. These include: the definition of content analysis and the evaluation of traditional content analysis. Section 2.6 highlights the importance of the computerised content analysis in general and *Nudist* in particular in the qualitative research. Section 2.7 focuses on the benefits of using *Nudist* software in the present study. Section 2.8 summarises the main issues discussed in the chapter.

2.2. The Importance of Annual Reports to Financial Analysts

Information is disclosed by firms in a number of ways. In addition to the annual report, there are a number of other sources that might provide investors with value-relevant information in predicting firms' future performance. These sources include: interim reports, press releases, conference calls and direct communication with analysts. To measure the quality of corporate disclosures, however, the present study focuses on the amount of forward-looking voluntary disclosure provided in its annual report narratives.

The decision to focus on the annual report is justified for a number of reasons. First, the annual report is a statutory document and it is required to be produced on an annual basis. Second, timing differences are minimized as most companies release their annual reports within three/four months after the financial year-end. Third, given their formalized structure, annual reports are more easily comparable among firms than other, less formal communication channels like press releases or direct contact with analysts. Fourth, the annual report is consistently ranked highly as a communication source by different groups of stakeholders (see the review in Chapter 2). Fifth, annual report disclosure scores are positively correlated with other media of financial communications (see Botosan, 1997 and Lang and Lundholm, 1993) suggesting that firms coordinate their overall disclosure policy. However, caution should be used in interpreting the results: to the same extent that several means of communication are

positively correlated, the marginal effect of voluntary annual report disclosures is likely to be overstated. Finally, for a technical reason, I use the annual report alone in this study because it is available on an electronic version for large numbers of UK companies on *Dialog* so I can score these reports automatically via *Nudist*. Other sources of information are not available electronically for large numbers of firms, but it is recognised that in practice investors are likely to use all sources of information to inform decisions about companies.

The main purpose of corporate annual reports is to provide information that is useful to investors and their advisors.[3] Many researchers, such as Day (1986), Schipper (1991) and Williams et al. (1996), argue that financial analysts are considered as one of the most important users of accounting information.

A number of studies have addressed the usefulness of annual reports to financial analysts. These studies can be divided into two categories. The first focuses on eliciting analysts' opinions regarding the types of information they use or would wish to use in making their decisions. The second focuses on studies based on content analysis of analysts' reports.

Different methodologies are applied in the first group of studies, including questionnaires, interviews, protocol analyses and experimental studies. A brief summary of these studies is listed in Table 2.1. Most of these studies conclude that the annual report is the most important source of information, that the income statement is more useful than other financial statements and that direct communication with management is an important source of information. Also, some of these studies focus on the importance of non-financial information, such as a company's market and

[3] The annual report also provides useful information for a wide range of other users. They include managers, unions, customers, debtors, creditors and employees.

competitive position and business segment data (e.g., Pankoff and Virgil, 1970 and SRI, 1987). Only Mear et al. (1987) conclude that balance sheet information is more important than income statement information. Epstein and Palepu's (1999) shows that financial analysts regard annual reports as an important source of information, particularly the management's discussion and analysis (MD&A) and the president's letter. Beattie and Pratt (2002) find that management discussion and analysis items are rated highly by professional and non-professional users. These results are consistent with Schipper (1991) who suggests that annual report narratives are the most important source of information for professional users such as financial analysts.

One limitation of all these studies is that they focus on what analysts say about the information they use in making their decisions. However, these studies do not actually investigate the type of information that financial analysts used in writing their reports. Also, it is difficult to generalise the results of these studies because the sample sizes used in some of them are relatively small. Finally, there is a possibility that the instrument (i.e., a questionnaire or an interview) used to collect analysts' opinions has adverse influences upon their response.[4]

The second group of studies, and the group most closely related to the present study, contains papers that use a content analysis approach. This approach is used to investigate the types of information that are actually used in sell-side financial analysts' reports. Using this approach helps to overcome the problems inherent in the

[4] More specifically, questionnaires, for example, are standardised. They contain a limited number and type of information items identified by the researcher. This constrains the analyst's choice of information items that they may find useful. By allowing frequent space for comments, however, the researcher can partially overcome this disadvantage. The second disadvantage of using questionnaires is that respondents may give superficial answers, especially if a long period of time is needed to answer many questions. There are also many disadvantages of using interviews. The most crucial one is that the findings of a study based on interviews are basically dependent on the researcher's skills in asking questions, in analysing the data and in ensuring confidentiality.

methodologies used in the previous studies, by focusing on the final output of financial analysts' work.

Table 2.1. Summary of Studies of the Usefulness of Accounting Information to Financial Analysts (In chronological order)

Author(s) and Year	Research objectives	Methodology	Findings
Pankoff and Virgil (1970)	Examining the usefulness of annual reports information and other information to financial analysts.	Experimental study.	Analysts find that non-financial information such as general economic and industries information is more important than financial statements information.
Chandra (1975)	Investigating the information needs by financial analysts.	Questionnaire.	Income statement in general and EPS in particular are more important than balance sheet information.
Horngren (1978)	Investigating the information needed by financial analysts.	1- Review analysts' reports. 2- Interviews.	1-Annual report in general and income statement in particular is the most important source of information. 2-Communication with managers is an important source of information.
Arnold and Moizer (1984)	Identifying the process used by financial analysts to appraise investments in ordinary shares.	1- Questionnaire. 2- Interviews.	1-Analysts use the following sources of information: Income statement, balance sheet and the interim results. 2-The most important sources of information are discussions with management and the chairman's statement.
Biggs (1984)	Examining the information used by financial analysts to assess corporate earnings power.	Protocol Analysis.	Analysts use income statement information, such as operating performance ratios, more frequently in assessing corporate earnings power.
Moizer and Arnold (1984)	Comparing the information used by sell -side and buy-side analysts.	1-Questionnaire. 2- Interviews.	Sell-side analysts consider the annual report, interim reports and discussions with managers as important source of information.

Chang and Most (1985)	Examining the information used by individual and institutional investors.	Questionnaire.	The annual report is the most important source of information.
Day (1986)	Studying the usefulness of the annual report to financial analysts, who followed the chosen company and those who do not.	1- Protocol Analysis. 2- Content Analysis.	1-Analysts use the annual report information more frequently. 2- Specialised analysts use other sources of information about future prospects such as industry statistics, preliminary figures and company contacts. In contrast, non-specialised analysts use past information such as previous accounts and share price data.
Bouwman et al. (1987)	Investigating the information used by analysts to make their decisions.	Protocol analysis.	Analysts frequently use income statement, segmented information, information about products and markets.
SRI FERF (1987)	Identifying the information needs of both individual and professional investors.	1- Questionnaire. 2- Interviews.	1- Analysts frequently use the annual report, but they ranked communication with management as the most important source of information. 2- Analysts are also interested in information about the company's market and competitive position, management goals and objectives, segmented information.
Mear and Firth (1987)	Assessing the importance of information to analysts when they make their decisions.	Experimental study.	Analysts use balance sheet information more frequently than income statement information. These include net assets, proprietorship ratio and liquidity.
Mear and Firth (1988)	Examining the relevance of the annual report information for risk assessment.	Experimental study	Analysts use balance sheet information such as net assets, proprietorship ratio and profitability ratios for assessing a company's risk.

Weetman et al. (1994)	Examining the views of analysts on the content of the Operating and Financial Review (OFR) statement.	Interviews.	1-The annual report is the most important source of information. 2- The OFR statement contains useful information to analysts in making their forecasts.
Brown (1997)	Determining the usefulness of the annual report and the interim report to sell-side analysts.	Questionnaire.	Annual reports and interim reports have useful information for financial analysts
Barker (1998)	Examining the economic incentives of finance directors, analysts and fund managers with respect to stock market information flows.	1-Observation. 2-Questionnaire. 3-Interviews.	Analysts' sources of information are direct contact with the company, analysts' meetings, interim and annual results announcements and other presentations for groups of analysts.
Mavrinac and Siesfeld (1998)	Examining the usefulness of non-financial information to analysts.	1- Questionnaire. 2-Experimental study.	Analysts rely on non-financial information when making their decisions. These include the execution and the quality of corporate strategy, management credibility and experience and market share.
Epstein and Palepu (1999)	Examining the usefulness of narrative disclosure in the annual report to financial analysts.	Questionnaire.	Financial analysts rank annual reports as an important source of information, particularly the MD&A and the president's letter.
Beattie and Pratt (2002)	Examining the annual report users' views about a comprehensive set of 130 disclosure items.	1- Questionnaire. 2- Follow-up telephone interviews.	Narrative disclosures such as disclosure of broad objectives and strategy, together with some of the management discussion and analysis items, background items, risk items and innovation value driver items are rated most highly by all users.

Content analysis has been used in the accounting and finance literature at least since 1976. This method is defined by Jones and Shoemaker (1994:142) as 'a research method, which draws inferences from data by systematically identifying characteristics

within the data'. The authors also provide a complete review of studies using this approach in accounting and finance between 1976 and 1993.

By examining analysts' reports, this method provides an opportunity to understand the types of information that are used by analysts in their decision-making process. Hence, one can evaluate the usefulness of corporate disclosure to one of the most important users of financial reports.

Academic research in the information content of analysts' reports include Horngren, 1978; Govindarajan, 1980; Previts et al., 1994; Bricker et al., 1995; Mavrinac and Boyle, 1996; Rogers and Grant, 1997; Breton and Taffler, 2001; Rogers and Fogarty, 2001; Bradshaw, 2002; and Abdolmohammadi et al., 2003. Of these, the following articles are the ones which most closely relate to the present study.

Rogers and Grant (1997) is the first study which investigates the link between the information published in annual reports and the information cited by financial analysts in their reports. Unlike Previts et al. (1994), their study identifies the potential sources of the information included in sell-side financial analysts' reports.

Rogers and Grant (1997) investigate a sample of 187 US sell-side financial analysts' reports on publicly quoted firms, whose financial years ended between July 1, 1993 and June 30, 1994. They use the 'sentence' as the information unit when coding the full text of each analyst report. They then classify each information unit into one of six categories: financial and operating data, analysis of financial and operating data, forward-looking information, management and shareholder information, firm description, and description of the firm's operating environment. Finally, they relate each category to the information included in annual reports.

The results of the Rogers and Grant (1997) study indicate that financial statements provide only about 26% of the information included in sell-side analysts' reports and that the income statement is considered as the most important source of information for financial analysts. In addition, they find that narratives in US annual reports are the most important source of information used by financial analysts. They find that the MD&A and the chairman's statement contain an additional 26% of the information cited in analysts' reports.[5] Surprisingly, around 48% of the information comes from external sources. These findings are consistent with the argument by Anandarajan et al. (2000), who state that current accounting practices force the users of the annual report to rely on other sources of information in order to learn whether a company's technologies, knowledge capital and procedures appear relevant to the market they serve.

The most crucial finding in Rogers and Grant (1997) is that narrative disclosures in the annual report, especially the MD&A section, are the most important source of information for financial analysts. A Significant portion of these discussion sections are cited by financial analysts in their reports.

Other studies also highlight the importance of narrative reporting to sell-side financial analysts. Breton and Taffler (2001), for example, find that UK financial analysts use information about management, company strategy and the company's trading environment in making their recommendations. Additionally, Mavrinac and Boyle (1996) investigate the extent to which sell-side analysts use non-financial information in making their decisions. They find that analysts following growth firms refer to customer

[5] Rogers and Grant (1997) find that annual report narratives provide, on average, 40 percent of the information cited in analysts' reports. However, they notice that the amount of common information provided by both narrative sections and the financial sentences is 14 percent. Therefore, their results indicate that narrative disclosures incrementally count for only 26% of the information cited in analysts' reports.

and activity information more frequently than analysts following slowly growing firms. Moreover, they report that analysts following growth firms refer to detailed information about new products, research and development and other innovation investments, while analysts following slowly growing firms concentrate on internal operations and manufacturing information. They also observe that analysts following service firms publish detailed information about employees, their training and the efficiency of service provisions.

A further group of studies examine the usefulness of particular parts of the annual report discussion section. For example, Steele (1982), Abrahamson and Amir (1996) and Smith and Taffler (2000) argue that the information content of the chairman's statement is useful with respect to the future performance of the reporting firm. Clarkson et al. (1994) and Bryan (1997) focus on the importance of a firm's MD&A disclosure, while Schleicher and Walker (1999) and Walker and Tsalta (2001) emphasise the importance of the Operating and Financial Review (OFR).

The literature cited above highlights the importance of the annual report discussion section. This importance can be attributed to '(1) the rapid pace of change in business, which means that past performance has become a less useful guide to future prospects and (2) the growth of intangible assets, including those generated from intellectual capital, that are not recognised in the traditional financial statements' (Beattie and Pratt, 2002). As mentioned before, the aim of the thesis is to examine the economic consequences of increased the level voluntary annual report narratives disclosures. More specifically, it examines the effect of forward-looking disclosures in annual report narratives on the market's ability to anticipate future earnings changes. Therefore, the following section provides further discussion relating to the importance of forward-

looking information. It starts with a definition of forward-looking information. Then, it discusses the arguments for and against this type of information. Finally, the studies that emphasise the economic benefits of this type of information are reviewed.

2.3. Forward-Looking Disclosure

2.3.1. Definition of Forward-Looking Information

Information published in the annual report can be classified as 'backward-looking information' and 'forward-looking information'. Backward-looking information refers to past financial results and their related disclosures. Forward-looking disclosure refers to information on current plans and future forecasts that enable shareholders and other investors to assess a company's future financial performance. Such forward-looking information involves, but is not limited to, anticipated operating results, anticipated financial resources, changes in revenues, cash flows and profitability. Forward-looking information also involves risks and uncertainties that could significantly affect actual results and cause them to differ from projected results. These risks and uncertainties include local, regional and national economic conditions, domestic and international political events, the effects of governmental regulation and the competitive environment in which the company operates. In many cases, one can identify forward-looking sentences by terms such as 'likely', 'will', 'forecast', 'expect', 'anticipate', 'estimate', 'predict,' or other comparable terminology.

The definition of the backward and forward-looking information is *not* as simple as stated above. This is because, in many cases, some types of information may be categorised as backward-looking while they carry messages which have relevance for the future. For example: if the CEO reports in the annual report that the level of R&D expenditure was increased by 10% last year. This sentence definitely refers to the past.

However, it implies that such investment in R&D is expected to lead to an increase in the future cash flow.

2.3.2. Arguments for and against Forward-Looking Information

There is a large literature that attempts to explain what motivates firms to voluntarily disclose additional information. Healy and Palepu (2001) and Walker (1997) provide comprehensive reviews of this literature. This section focuses only on one type of discretionary disclosure, forward-looking information.

There are various arguments concerning the merits of publishing forward-looking information. For example, Keiso and Weygandt (1995) present these arguments: (1) forward-looking information will be helpful to investors in their investment decision-making process, (2) the absence of forward-looking information may lead investors to base their forecasts on inaccurate information from other sources (e.g., inaccurate analyst forecasts), and (3) the economic environment is too dynamic to rely on historical information only.

In addition, Bujaki et al. (1999) argue that the publication of forward-looking information in the annual report is useful for reducing the degree of information asymmetry between managers and investors, thereby reducing the firm's cost of external financing. This argument is consistent with the capital markets transactions hypothesis as a motivation for voluntary disclosure (Healy and Palepu, 2001).

On the other hand, there are arguments against the inclusion of forward-looking disclosures. First, because of the uncertainty associated with the future, it might be difficult to predict with accuracy. In addition, firms might leverage their performance towards the level of their forecasts (Kasznik, 1999). Second, inaccurate forecasts might lead to lawsuits. This is consistent with the litigation cost hypothesis (e.g. Field et al.

2003). Litigation might reduce a manager's incentives to provide forward-looking information. This is especially true when managers believe that the legal system cannot distinguish between forecast errors due to uncertainty and deliberate management bias. Third, forward-looking disclosure might provide useful information to competitors and hence, might affect its competitive position in product markets. This is consistent with the proprietary cost hypothesis (Healy and Palepu, 2001).

2.3.3. The Nature of Forward-Looking Information

Studies that look at how forward-looking information is presented in the corporate annual report show that this type of information can be qualitative, quantitative, financial or non-financial. For example, Bujaki et al. (1999) use the content analysis methodology to describe the nature of forward-looking information published in the chairman's statement and the MD&A for 46 Canadian companies. They find that 19.2% of information included in the chairman's statement and the MD&A is forward-looking. In addition, they observe that most of the forward-looking information is qualitative and company-specific. Another important finding in Bujaki et al. (1999) is that good news dominates bad news. Good news disclosures account for 97.5%, while 2.5% of forward-looking information is bad news. This argument is broadly consistent with the findings in Clarkson et al. (1992 and 1994) and Clatworthy and Jones (2003).

Clarkson et al. (1992 and 1994) argue that managers tend to publish favourable forward-looking information in their annual reports. The findings in Clatworthy and Jones (2003) suggest that UK companies prefer to report positive aspects of their performance. In addition, their study shows that UK companies prefer to take credit for good performance themselves whilst attributing bad performance to external sources.

In addition to the above studies, large numbers of studies empirically investigate the economic benefits of disclosing forward-looking information. These studies are reviewed below.

2.3.4. The Economic Benefits of Forward-Looking Information

Numerous studies examine the benefits of forward-looking information in a variety of contexts. These include the prediction of corporate future performance, the characteristics of analyst forecasts and stock price behaviour.

A number of studies investigate the usefulness of forward-looking information for anticipating corporate future performance. One such study is Clarkson, Kao and Richardson (1994). Clarkson et al. (1994) find that the inclusion of forward-looking information in corporate annual reports is informative with respect to corporate future performance. They also notice that the decision to include this type of information depends upon managers' concern regarding the reaction of capital and product markets. They demonstrate that firms with good news are more likely to publish forward-looking information if financial market considerations dominate product market considerations.[6]

Another study that links corporate disclosure with corporate future performance is Bryan (1997). Bryan (1997) analyses a sample of 250 MD&As. The findings suggest that indications of envisaged future operations and capital expenditures are associated with future short-term performance measures, after controlling for information contained in financial ratios. Moreover, Clarkson et al. (1999) provide evidence that changes in the level of forward-looking information in the MD&A vary directly with

[6] The authors use different proxies to capture financial market considerations. These proxies are equity as a percentage of total assets and total debt and equity as a percentage of total assets. Furthermore, they interpret product market considerations in terms of barriers to entry. They use three different measures to proxy for this concept. These measures are capital requirement, the variability of the firm's return on equity and the four-firm concentration ratio.

future corporate performance. This suggests that forward-looking disclosures in the MD&A provide credible information.

Besides studies that focus on corporate future performance, there are studies that concentrate on the characteristics of financial analyst forecasts. For example, Barron et al. (1999) examine the association between corporate disclosure quality and the accuracy of analyst forecasts. They show that higher MD&A disclosure scores are associated with more accurate analyst forecasts. This result is driven by forward-looking information about capital expenditure and operations. In addition, Walker and Tsalta (2001) find a positive association between analyst following and the quality of forward-looking information published in UK annual reports.

A final group of studies examines the effects of increasing the level of corporate disclosures on stock prices. Miller and Piotroski (2000), for example, use the *Dow Jones News Retrieval* database to investigate the effect of the disclosure of forward-looking earnings information on stock prices.[7] Their results show that stock returns around the disclosure of this type of information are more correlated with the next period's earnings than those of firms without forward-looking disclosures.

Frost (2002) finds that the stock market responses to forward-looking information are greater than their responses to earnings and sales results. More specifically, he shows that the stock market responds significantly to 'soft' prospective disclosures, especially to earnings forecasts.[8] These results indicate that the announcement of forward-looking statements is highly informative in that they cause significant share price revisions.

[7] The authors use the *Dow Jones News Retrieval Service* to search the Dow Jones News Service, the PR Newswire, the Business Wire and the Wall Street Journal for forward-looking statements by management regarding earnings or other measures of performance including sales, cash flows, EBIT and margins.
[8] Frost (2002) refers to non-quantitative forward-looking disclosures as 'soft' information.

Hutton et al. (2003) also look at press releases and find that good news earnings forecasts are uninformative when issued without supplemental disclosures. In contrast, these forecasts become informative when managers supplement their earnings forecasts with 'verifiable' forward-looking statements regarding income statement line items. These results suggest that managers make 'verifiable' forward-looking statements to strengthen the credibility of good news earnings forecasts.

Finally, based on a sample of UK companies, Schleicher and Walker (1999) provide evidence that high levels of disclosure in the OFR statement enable the market to better anticipate future earnings changes. This effect is significant when the authors use a forward-looking disclosure index and when they use models that examine one-period-ahead and two-period-ahead share price anticipation of earnings.

The stream of research discussed above suggests that forward-looking disclosures are valuable to investors because they contain incremental information. This information is relevant in forecasting future performance. The evidence also suggests that narrative disclosures carry valuable information for financial analysts. Finally, the review above highlights the importance of forward-looking information to stock returns.

The present study tries to shed further light on the importance of forward-looking information voluntarily disclosed in UK annual reports. This is done by examining the impact of these disclosures on the informativeness of stock prices for future earnings changes. In order to achieve this, I construct disclosure indices to proxy for the quality of forward-looking information in annual reports. Disclosure scores are then linked with prices leading earnings. Therefore, the following section reviews the literature concerned with the identification of proxies for the quality of corporate disclosures.

2.4. Proxies for Disclosure Quality

A number of definitions of 'disclosure quality' are given in the literature. For example, Diamond and Verrecchia (1991) define disclosure quality as the accuracy of investors' beliefs about stock prices following the disclosure. King (1996) defines disclosure quality as the degree of self-interested bias in corporate disclosure. Hopkins (1996) defines disclosure quality as the extent to which current and potential investors can read and interpret the information easily.[9]

It is difficult to measure investors' perception of the firm's disclosure quality. Hence, different proxies are used in the literature. In a more recent survey of the literature, Healy and Palepu (2001) review studies that consider different proxies for the quality of corporate disclosures. They categorise these proxies into three groups: management forecasts, subjective ratings and self-constructed indices. The purpose of this section is to critically evaluate the advantages and disadvantages of these proxies.

2.4.1. Management Forecasts

Management forecasts are frequently used by US researchers.[10] *First Call* Database provides these forecasts for large numbers of firms. As a result, a large body of US studies uses these forecasts as a proxy for disclosure quality (e.g., Miller and Piotroski, 2000). The reasons for using these forecasts include the way that they can be accurately measured. This is due to the fact that managers sometimes provide either point or range estimates for earnings or revenues. Also, the exact period of time for these forecasts is typically identified. Consequently, researchers can measure variables such as management forecast accuracy.

---

[9] All these definitions are cited in Beattie et al. (2001).

[10] In the UK there is no database containing quantitative management forecasts. Using management earnings forecasts in disclosure studies may be an area for future research in the UK (see Chapter 8).

On the other hand, it is clear that management earnings forecast is only one component of managers' voluntary disclosure package. Companies voluntarily publish different types of information in their annual reports. For that reason, it is often not sensible to use this type of information only to proxy for the overall level of corporate disclosure quality.

Overall, using management forecasts as a proxy for corporate disclosure in market-based accounting research can be expected to enhance the power of the regression analysis. On the other hand, such forecasts cannot necessarily be used to represent the quality of the firms' overall disclosure quality.

2.4.2. Subjective Ratings

Subjective ratings provide aggregate measures of corporate disclosure quality. The most common ratings are the US ratings provided by the annual survey of the AIMR-FAF for the years 1980 to 1995. Below I provide a brief review of AIMR-FAF ratings and some other subjective ratings that are used in the literature.[11]

AIMR-FAF ratings

Every year a number of financial analysts meet in industry-specific subcommittees to evaluate the disclosure quality of firms relative to their industry peers. Evaluating disclosure quality is based on three media: disclosures in annual reports and required published materials, disclosures in quarterly reports and other non-required published materials, and information provided through investors' relations programs. The overall measure of a firm's disclosure quality is calculated as a weighted average of the three individual category ratings. AIMR-FAF ratings have been used in several studies, such

[11] Details on AIMR-FAF disclosure ratings are given in Lang and Lundholm (1993, 1996), Healy and Palepu (2001) and Core (2001).

as Lang and Lundholm (1993, 1996), Lundholm and Myers (2002) and Geld and Zarowin (2002).

The advantages of AIMR-FAF ratings can be summarized into three main points: (1) the ratings cover all the various disclosures made by firms, including verbal information given during analysts meetings and conference calls, (2) experts from brokerage firms provide these ratings. Hence, they are likely to be a reasonable proxy for corporate disclosure quality. This is because financial analysts are able to assess the information needs of the industry and they are familiar with the disclosed information and its use during the year and (3) errors in the judgment of individual analysts are diversified away when aggregating individual scores into consensus ratings.

On the other hand, AIMR-FAF ratings potentially have some serious problems. These include: (1) the ratings are not necessarily based on examining the original corporate reports (see Lang and Lundholm, 1993). Therefore, analysts' opinions may be subjective and inconsistent, (2) as discussed in Healy and Palepu (2001), it is not clear whether financial analysts take these ratings seriously and how they select firms to be included in the ratings. It appears that AIMR-FAF ratings are biased towards the largest firms in each industry and (3) AIMR-FAF ratings were discontinued in 1997 with the last year of the disclosure scores being 1995. Thus, these ratings become increasingly dated.

Other subjective ratings

Researchers use a number of other subjective ratings to proxy for the quality of corporate disclosures. Examples of such ratings include Financial Post ratings (Sutley, 1994), Australian Stock Exchange ratings (Brown et al., 1999), SEC ratings (Barron et al., 1999), Society of Management Accountants of Canada (SMAC) ratings (Richardson

and Welker, 2001), Actualidad Economica ratings (Blasco and Trombetta, 2002) and Center for International Financial Analysis and Research (CIFAR) ratings (Hope, 2003a, b and c). In all these studies, it is not clear whether these ratings are based on investigating the publications of firms or whether they just reflect analysts' and accountants' general opinions regarding the firm's disclosure policy.

One way to mitigate the inherent problems in subjective ratings is to use self-constructed disclosure indices that are based on a list of disclosure items in evaluating the quality of corporate disclosure.

2.4.3. Self-Constructed Disclosure Indices

The concept of disclosure quality is very difficult to assess. This is because this concept refers to the extent to which current and potential investors can read and interpret the information easily (see Hopkins, 1996). Measuring investors' perception of the firm's disclosure quality is not an easy task. In addition, it is difficult to measure qualitative attributes of the annual report or other disclosures such as reliability, diagnostic, value and the interaction between different report items. As a result, researchers tend to assume that the *amount* of disclosure on specific topics proxies for the quality of disclosure (see Beattie et al., 2002). In addition, disclosure index studies assume that the quantity of corporate disclosure and the quality of corporate disclosure are positively related (Botosan, 1997).

Cerf (1961) seems to have been the first study to evaluate the extent of corporate disclosure through the construction of a disclosure index. Since Cerf (1961), researchers have extensively used disclosure indices to assess, compare and explain differences in the amount of information disclosed in corporate annual reports. Marston and Shrives (1991), Schleicher (1996) and Schadewitz and Blevins (1997) provide comprehensive

reviews of the use of disclosure indices in evaluating the quality of corporate disclosures.

Self-constructed disclosure indices can be classified into two groups. The first group reflects more or less all items that are published in annual reports. However, these studies do not match the items in the disclosure index to a specific user group (see Cooke, 1989a and b, 1991; Schleicher and Walker, 1999). The second group links the included disclosure items to specific user groups (see Wallace, 1988).

Another classification is provided in Beattie et al. (2001). The authors classify self-constructed disclosure studies as a partial content analysis and a holistic content analysis. In a partial content analysis, researchers identify a list of disclosure topics. They then text-search the annual report for the presence of these topics. In a holistic content, analysis researchers investigate the whole annual report to construct their disclosure index.

The construction of a disclosure index typically involves three stages:

Selecting the preliminary list of disclosure topics

The first stage is to select the items of information which might appear in the annual report. Wallace and Nasser (1995) state that there is no general theory on the items to be selected for investigating the extent of disclosure. The selection of the preliminary list of disclosure topics is usually based on reviewing the literature and on reading a sample of corporate annual reports.

Including large numbers of disclosure topics in a preliminary list can assist the users of financial reports in selecting the most relevant disclosure topics. This decreases the subjectivity and bias that would be involved if the researcher presents only a few disclosure topics that represent his/her own opinion regarding the topics that should be

disclosed. On the other hand, the inclusion of too many topics in a preliminary list is a concern if financial analysts, for example, perceive the task of weighting these items as too big to give it focused attention (Hooks et al., 2000).

Selecting the final list of disclosure topics

Selecting the final list of disclosure topics is the most crucial stage in determining the final ranking of annual reports. When constructing that list, most disclosure studies explicitly refer to a particular user group. The resultant list is then restricted to items which are likely to be relevant to that particular user group.

Different methodologies are applied in the literature to select the final list of disclosure topics. These methodologies include: (1) sending out questionnaires to the users of financial reports, (2) conducting interviews or (3) relating to recommendations provided by the accounting profession and accounting standards.

Measuring the quality of disclosure

Once the final list of topics is identified, the researcher can use this list to create a disclosure score for each firm. First, an appropriate score is allocated to the firm if its annual report contains a particular piece of information. These scores represent the partial scores. Second, individual scores are then aggregated into a total index which summarises the overall quality in a single number.

Disclosure indices normally measure the extent of corporate disclosure by allocating a score of 1 for the presence of disclosure topics and a score of 0 otherwise. In some studies (e.g., Firth, 1980), disclosure topics are weighted for their relative importance. The crucial point is that the choice between these different weighting methodologies frequently does not significantly change the research results (Schleicher, 1996). Chow and Wong-Boren (1987) and McNally et al. (1982), for example, find that the level of

voluntary disclosure increases with firm size. Their result is obtained regardless of whether they use the importance-weighted or equally-weighted index as the independent variable. This appears to indicate that 'firms which are better at disclosing important items are also better at disclosing less important items and therefore are consistent in their disclosure policy' (Schleicher, 1996:53).

Corporate disclosures are not easy to evaluate because the construction of a disclosure index requires subjective assessments by the researcher(s). As a result, it is essential to assess the reliability and the validity of the disclosure measure. The following two subsections discuss the definition of these concepts. They also provide a detailed discussion regarding the way to measure these concepts.

2.4.3.1. Assessing the Reliability of a Disclosure Index

A central problem of content analysis is related to the data reduction stage when the whole text of a report is classified into a much smaller set of content categories. Weber (1990:12) argues that 'to make valid inferences from the text, it is important that the classification procedure be reliable in the sense of being consistent: different people code the same text in the same way'. In addition, Weber (1990:15) notes that 'reliability problems [are] usually caused by the ambiguity of word meanings, the ambiguity of category definitions or other coding rules'. Krippendorff (1980) argues that there are two forms of reliability that must be assessed within content analysis: stability and reproducibility.

Stability is concerned with the extent to which the results of content classification are stable over time. Stability can be established when the same text is coded more than once by the same researcher. Inconsistencies in coding lead to unreliability. These inconsistencies can arise if the coding instruments are ambiguous, cognitive changes

within the researcher or simple errors such as recording the wrong text unit for a category. Stability is considered the weakest form of reliability. This is due to the fact that only one researcher codes the text.

The second type of reliability is reproducibility, also known as inter-reliability. Reproducibility is related to the extent to which content classifications yield the same results when the same text is coded by more than one researcher. Conflicting coding usually results from cognitive differences among the researchers, ambiguous coding instructions or errors. The main difference between stability and reproducibility is the fact that stability measures the consistency of one researcher's understandings, whereas reproducibility measures the consistency of different researchers' understandings.

A common way to achieve inter-reliability is by using several researchers and ensuring that differences between the researchers are few or that differences have been re-analysed and then resolved. Also, developing a set of explicit and well-specified instruments or using computer software helps to increase reliability.

2.4.3.2. Assessing the Validity of a Disclosure Index

The second problem deals with the validity of variables used to identify the content classifications. Validity is referred to as the extent to which a study accurately reflects or assesses the specific concept that the researcher is attempting to measure. It is argued that content analysis itself is only valid to the extent that the results are related to other measures (Shapiro and Markoff, 1997).[12] O'Connor (2003) suggests three methods of assessing validity. These are face validity, content validity and construct validity.

[12] For example, Botosan (1997) supports the validity of her disclosure index by four analyses. These are: (1) the correlation among the components of disclosure index, number of analysts and the number of Wall Street Journal articles, (2) Cronbach's coefficient alpha, (3) the correlation between disclosure scores and firm characteristics identified in prior studies to be associated with the level of corporate disclosures, and (4) the correlation between the disclosure scores and the AIMR-FAF disclosure scores.

Face validity is an assertion that the researcher reasonably measures what he/she intends to measure. Usually, a researcher can ask a colleague or an expert to confirm that the classifications measure what the researcher intends to measure.

Content validity is an assertion that the researcher covers most of the classification items. There are different ways to estimate this but one way in the disclosure indices studies is to calculate the correlation between disclosure scores calculated by a labour–intensive approach and a computerised approach. The current study presents this comparison later.

Construct validity is related to the extent to which a classification scheme is thoroughly grounded in the theory and empirically used by different researchers. For example, it is possible to support the validity of disclosure scores by finding a significantly positive correlation between corporate disclosure scores and firm characteristics identified in prior literature to be associated with the levels of corporate disclosures (see Footnote 12).

2.5. Content Analysis

2.5.1. Definitions

There are many definitions of content analysis. The most widely accepted definition is the one offered by Berelson (1952). Berelson (1952:18) defines content analysis as 'a research technique for the objective, systematic and quantitative description of the manifest content of communication'. Similarly, Carney (1972:21) describes it as 'a research technique for making inferences by objectively and systematically identifying specified characteristics of messages'. Krippendorff (1980:21) describes this approach as 'a research technique for making replicable and valid inferences from data to their context'. Other definitions are provided in Weber (1990) and Holsti (1969).

2.5.2. Evaluation

Content analysis has many advantages. First, it looks directly at texts and, hence, looks at a central aspect of social communication. Second, it is easy to perform, and it allows both quantitative and qualitative analyses. Third, it provides valuable historical insights through the analysis of textual content for different periods. As a result, it is useful for examining trends and patterns in documents over time.

Traditional content analysis via manual reading is generally more effective than the computerised analysis in identifying certain themes in the texts. But the traditional approach has been the subject of many criticisms. In particular, this approach can be extremely time-consuming and expensive, specifically when one analyses large volumes of textual data. Also, human coders could make mistakes during their analyses. They could overlook some text of relevant content, potentially affecting the validity of the measure.

One way to reduce the limitations of the traditional content analysis is to use the computer-based content analysis. The next section provides detailed discussion regarding this approach.

2.6. Computerised Content Analysis

2.6.1. Overview

Increasingly, researchers in various disciplines are using computer software packages to identify and access content, to create content categories and to analyse the occurrence and frequency of terms in context.

Holsti (1969) suggests four cases in which the use of computerised content analysis might be useful. These are: (1) when the unit of analysis is a word and the analysis is concerned with how many times a word is used, (2) when the analysis is extremely

complex because of a large number of texts or a large number of categories, (3) when the analysis involves analysing the data in multiple ways, and (4) when the data is of basic importance to a variety of disciplines and might be used in multiple studies.

Holsti (1969) also suggests another four cases in which the use of computers might not be suitable. These are: (1) when the research involves a single expensive specialised study, (2) when the number of documents is large but the information is limited, (3) when the research calls for measures of space or time, and (4) when thematic analysis is being used.

The computer-based content analysis has several advantages over the traditional content analysis. If the materials are archived in an electronic database, then the computer program can provide an organised storage 'file' system, so that the researcher can quickly and easily store text files in one place. The computer program helps a researcher to locate the data easily, whether this data is an idea, a statement, a phrase, or a word. Also, the search for text can be easily accomplished with a computer program. This helps a researcher to avoid using manual archives to make copies of the texts.

Furthermore, electronic archives are designed to provide advanced searching capabilities to help identify relevant content. Computers minimise the time needed for routine counting. They might also produce high levels of reliability, assure greater degrees of reproducibility and reduce the problems of researcher bias.

Finally, the computer-based content analysis is more flexible than the traditional content analysis. At any time during the coding process, decisions can be changed. In such cases computer software packages allow the researcher to easily try different options and to assess the impact of any new decisions on the resulting output.

2.6.2. The Use of Nudist

Nudist is a computer program designed to aid users in handling non-numerical and unstructured data in a qualitative analysis. In the present study, *Nudist* is used to assist in the identification of the disclosure items from analysts' reports. It is also used to *automatically* score large samples of annual report narratives.

Recently, *Nudist* has been used in accounting studies to explore the type of information published in UK annual reports (see Beattie et al., 2002), US annual reports (see, Schnatterly, 2003) and US analysts' reports (see Rogers and Fogarty, 2001).[13]

The present study has several of the attributes that Holsti (1969) mentions for a successful application of computerised content analysis. In particular, the analysis involves large text files easily accessible in an electronic format and the data is analysed in multiple ways. The main benefits of using *Nudist* in my study include: (1) the ability to store and organise files, (2) the ability to parse the data down to a manageable size, (3) the ability to identify and index every text unit in the file as an individual object for the purpose of analysis, (4) the ability to search the text units for specific words and variants of those words, (5) the ability to count the frequency of given key words, (6) the ability to score very large numbers of text files at low marginal costs, (7) the ability to compare disclosure scores across firms and over time, (8) the ability to combine the results of words searches using Boolean operators such as 'AND' and 'OR', (9) the ability to export the results into a spreadsheet and then into a statistical analysis

[13] A number of disclosure studies use different computer software packages. These include *WORDS* (see Frazier et al. (1984); *WordCruncher* (see, e.g., Previts et al., 1994, Bricher et al., 1995, Previts and Robinson, 1996 and Rogers and Grant, 1997); *WORDS*, *OCP* and *SPSS-X* (see Smith and Taffler, 2000), *SATO* (see Breton and Taffler, 2001) and *General Inquirer* (see Kothari and Short, 2003). In addition, Abrahamson and Amir (1996) use their own developed computer program and Clatworthy and Jones (2003) use a computer program however they did not mention to its name.

package, like SAS or SPSS, and (10) the ability to replicate the scores easily in subsequent disclosure studies.

Nudist has three major functions for analysing textual data. These are text search, node search and command files. The nature of each function is described below.[14]

Text search

Nudist allows searching for the frequency of a single key word. However, unlike word processors that allow the researcher to search only one text file at a time, *Nudist* allows searching for a single key word across several text files at the same time.

The text-search results in a new node including all identified text units. For example, if one wishes to code any sentence referring to 'research and development', one way to do this is to text-search the document for this key word. Then *Nudist* automatically collects all sentences referring to this key word in one category called 'node'.

In addition, *Nudist* provides a report containing all text units for each searched key word. In the report, one can see the searched key word, the name of the document and the text units associated with the searched key word. The example on page 57 shows the results of a search for the key word 'research and development' in five annual reports where company names are replaced with their *Datastream* codes. Only the last firm has two text units including the key word. Each text unit is linked with a number. This number represents the location of the text unit within the annual report. A statistical summary is provided for each firm, which includes the percentage of the identified text units out of the total number of text units in the whole annual report. For example, the statistic summary for the last firm in the example is 2 text units out of 229. Finally, an

[14] Details on the *Nudist* commands, dialogs and functions are given in Richards (2000a) and Richard (2000b).

overall summary is provided for the whole sample. This includes the total number of text units found, percentage of reports that are associated with the key word, and the total number of text units in the whole sample.

Example: Report created by *Nudist* for a text-search of five annual reports for the key word ‘research and development’.

```
QSR N5 Full version, revision 5.0.
Licensee: Manchester University
PROJECT: Example, User Khaled Hussainey, 11:11 pm, Dec 13, 2003.
++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
+++ Text search for 'research and development'

+++ Searching document 135084...

+++ Searching document 135090...

+++ Searching document 135116...

+++ Searching document 135229...

+++ Searching document 135869...
RESEARCH AND DEVELOPMENT expenditure for the year again increased by
6%.                                                    36
We carry out primary RESEARCH AND DEVELOPMENT with ownership of the
intellectual properly,     resulting in unique and patent-protected technologies with real
longevity.                                    77
+++ 2 text units out of 229, = 0.87%
++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
+++ Results of text search for 'research and development':
++ Total number of text units found = 2
++ Finds in 1 documents out of 5 online documents, = 20%.
++ The online documents with finds have a total of 229 text units,
   so text units found in these documents = 0.87%.
++ The selected online documents have a total of 1731 text units,
   so text units found in these documents = 0.12%.
++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
```

Text searches in *Nudist* have four basic forms. The nature of each form is described below.

String searches

String searches are used to find specific words and phrases in the text files. When running a string search, *Nudist* will find all text units that contain identical occurrences of the string, whether the word stands alone or is embedded in a larger expression. For

example, if one searches for the word 'estimate' *Nudist* will find text units such as ‘estimate’, ‘estimates’, ‘estimated’, ‘underestimate’, ‘underestimates’, ‘underestimated’, ‘overestimate’, ‘overestimates’, and ‘overestimated’. If the researcher wants to restrict the search only to the word 'estimate', then it is necessary to select the ‘Whole Word’ function.

Pattern searches

Pattern searches find words with similar meanings or with different tenses. For example, occurrences of any of the words: ‘forecast’, ‘forecasts’, ‘is forecasted’, ‘are forecasted’, ‘is forecasting’ and ‘are forecasting’ are found by typing the pattern ‘[forecast|forecasts|is forecasted|are forecasts|is forecasting|are forecasting]’. Furthermore, this search type finds occurrences of words that have a common stem. For example, the words ‘profit’, ‘profits’ and ‘profitability’ are found by choosing the pattern ‘[profit|profits|profitability]’.

Case sensitive searches

This refers to the ability of *Nudist* to find all occurrences of a single key word exactly as the researcher types it in. For example, if the researcher types a key word in lower case, any occurrence at the start of a sentence will be excluded from the results.

Whole word or phrase only searches

In this case the search is restricted to the exact key word. The search will not find any key word embedded in a larger concept or expression. For example, if the researcher searches for the word ‘will’, *Nudist* will find all sentences that include this exact word ‘will’. The word ‘goodwill’, for example, will not be included in the results.

Node search

Node searches offer a unique tool for cross-examination of data via coding with a logically complete set of ways of asking questions about the researcher's codes and coding (Richards, 2000a).

There are seventeen Boolean search options that allow the researcher to perform more complex searches in *Nudist*. The most important options include intersection, union, overlap, matrix and vector searches.

The intersection search collects only text units that are common to the named nodes in a new node. In the present study, this search function is used to collect all coded sentences with a relevant topic, only if they are also coded as forward-looking sentences. A union search collects all the text units from the named nodes in a new node, while the overlap search finds all text units coded at any of the named nodes if they have at least one text unit in common.

A useful feature of *Nudist* is its ability to export the main results to a spreadsheet through using the matrix and vector search features. These features assist the researcher to obtain the intersections between different nodes in one stage, thus eliminating the need to perform repeated tasks to accomplish the same goal.

Command files

Using Command files allows one to analyse large samples of text files very quickly. A command file is a short instruction that functions much like a macro to automate repeated tasks, such as importing and text-searching a large number of documents.

2.7. The Present Study

In evaluating disclosure index studies, one can observe the following limitations: First, most of these studies are based on the information included in annual reports. Although

the annual report is the most important source of information to various users, there are other disclosure channels employed by firms in their communication with different users. These channels include presentation to financial analysts, interim reports and conference calls. Second, the methodology adopted in the previous literature has the following drawbacks:

1- The methodology can be characterised as a time-consuming process. One can spend a long time in selecting the information items in corporate annual reports and to score a sample of these reports.

2- Partial content analysis (where a researcher includes a few topics that represent his/her opinion regarding the items that should be disclosed) involves subjectivity and bias. Hence, it constrains the analysts' ability to choose the information items that may be useful in their decision-making process.

3- The process of ranking the importance of the pre-prepared disclosure items by the users of financial reports reflects their opinions about the information disclosed in financial reports, not the actual types of information they used in making their decisions.

The present study uses self-constructed disclosure indices to measure the quality of corporate disclosure.[15] However, it uses a new methodology to select the list of disclosure topics.

An important feature of this methodology is that it uses the textual content of analysts' reports to select the final list of disclosure items. It focuses on the types of information that financial analysts actually refer to in their reports. Thus, it avoids any potential research bias.

[15] Following the discussions in Botosan (1997) and Beattie et al. (2002), the present study assumes that the *amount* of disclosure on specific disclosure items proxies for the quality of corporate disclosures.

The rationale behind using analysts' reports to select the disclosure items can be summarised as follows. First, it is intended to discover the topics that help the stock market to anticipate future earnings changes. Since the 'market' is unobservable, it is assumed that financial analysts' reports can be used as a proxy for the market's view about the firm's disclosure quality.[16] This is because financial analysts are widely considered to be a very influential user group in stock markets. They represent and influence investors' beliefs and activities (e.g., Schipper, 1991; Hirst et al., 1995; and Lang and Lundholm, 1996). The final output of their work is written in reports that are made available to interested investors. Second, it is also known that financial analysts use a variety of different sources of information in writing their reports such as the annual report, personal meeting with managers and conference calls. Therefore, examining analysts' reports provides the basis for a comprehensive scoring sheet. Such a sheet should reflect the information used by financial analysts which are relevant for their decision making process.

Another important feature of the methodology in the present study is that it uses *Nudist* software to identify the list of disclosure items. A requirement of using *Nudist* for the above purpose is the availability of documents in an electronic format and the availability of a list of key words. As mentioned earlier, analysts' reports are used to construct the list of disclosure items. Analysts' reports are available in an electronic format from the *Investext Plus* database. Additionally, a list of forward-looking key words is developed in Chapter 4.

I use both traditional and computerised content analysis methods to select the list of disclosure items. A computerised method assists in selecting forward-looking sentences

[16] This assumption is motivated by the use of AIMR-FAF ratings as a proxy for the quality of corporate disclosure.

and in counting the frequency of forward-looking key words in each report. In addition, the traditional method is used by reading forward-looking sentences to identify the key topic in each sentence.

The identification of disclosure topics is basically carried out in three stages. The first stage involves creating a list of forward-looking key words. In the second stage, this list is used to identify forward-looking sentences that appear important to analysts in assessing a firm's future. Finally, each forward-looking sentence is read in order to identify the main topic of the discussion.

2.8. Summary

The present chapter reviews the literature relating to the importance of narrative disclosures. Researchers have focused on the importance of different aspects of corporate disclosure, such as the MD&A in the USA and the OFR in the UK. They argue that these narrative sections are value-relevant to the market. This literature justifies my decision to focus on this type of information in the present study.

In prior literature, researchers use different proxies for the quality of corporate disclosures. These are management forecasts, subjective ratings and self-constructed indices. The advantages and disadvantages of each proxy are presented in this chapter.

The present study uses a structured content analysis of the textual content of analysts' reports to produce a list of disclosure items. It uses the *Nudist* software to assist in the identification of the disclosure items. This chapter also highlights the advantages of using *Nudist*.

Chapter 3: Share Price Anticipation of Earnings and the Quality of Corporate Disclosure: A Review

3.1. Overview

The aim of the present study is to identify the economic consequences of enhanced voluntary disclosure in the UK annual report discussion section. The study seeks to provide evidence that the level of forward-looking disclosure in corporate annual reports is significantly associated with the ability of investors to anticipate future earnings changes.

The present chapter reviews the existing literature relevant to the empirical analysis in the thesis. It focuses on two streams of literature. The first is related to the association between stock returns and earnings. The second is related to the economic consequences of corporate disclosure. Therefore, the aim of this chapter is to review existing studies in these two areas and to highlight certain gaps in the previous literature.

To date there has been relatively little attention given to the economic consequences of corporate disclosure in the UK. In contrast, academic research is quite extensive in the US. This is partly due to the availability of the AIMR-FAF ratings in the US. In the UK, there are no ratings similar to the AIMR-FAF ratings available. Therefore, the opportunity to undertake large-scale disclosure studies is limited.

The structure of the present chapter is as follows. Section 3.2 begins with a review of empirical studies on prices leading earnings. This is followed by a discussion of the Collins et al. (1994) paper. Section 3.3 surveys the literature regarding the economic consequences of corporate disclosure. The key articles relating to the effect of corporate disclosure on prices leading earnings are given special attention. Section 3.4 highlights the gaps in the existing literature and explains why the present study can make a valuable contribution for filling these gaps. Section 3.5 summarises the chapter.

3.2. Prices Leading Earnings

In the present section, return-earnings regression models that allow earnings to be anticipated some periods ahead are reviewed. The literature on the return-earnings association has undergone major developments since Ball and Brown (1968).[17]

Reported earnings have attracted major interest in the MBAR literature since the mid-1960s. The MBAR literature emerged as a response to the market efficiency hypothesis. Market efficiency is concerned with the extent to which security prices fully reflect all available information. Since accounting is an important source of value-relevant information about companies, it was natural for researchers to examine the efficiency of the market with respect to accounting information.

Ball and Brown (1968) analyse the price-earnings relation over a one-year return period. The one-year return period corresponds to the period covered in annual reports. They find that only 10 to 15 percent of the earnings announcement is new information to the market.

Following Ball and Brown (1968), many subsequent papers investigate aspects of the relation between stock returns and earnings. Lev's (1989) study provides an important review of empirical return-earnings studies between 1968 and 1989. Lev (1989) notes that the R^2s obtained by regressing annual stock returns on annual earnings or earnings changes are very low. They rarely exceed 10%. The author ascribes the weak explanatory power to the low quality of accounting earnings. The study of Lev (1989) challenges researchers to identify the potential explanations for the weak contemporaneous return-earnings association.

[17] These developments are summarised in three review articles: Cho and Jung (1991), Brown (2001) and Kothari (2001)

There are at least four potential reasons for the weak contemporaneous return-earnings association. These are: (1) market inefficiency, (2) stock markets react to value-relevant information that is not observed by the company, (3) noise in reported earnings, and (4) earnings may lack timeliness.[18]

The first potential reason for the weak return-earnings association refers to the market's informational inefficiency. This concept relates to the extent to which the stock market responses inappropriately to the reported earnings numbers. This phenomenon refers to the extent to which the stock market is able to take into account the changes in accounting methods when valuing companies. Walker (2004) clarifies this phenomenon by the following example. He argues that in calculating the reported earnings one of the judgements to be made is the selection of inventory valuation method. If the stock market is rational, such selection will not affect the market value of the firm. Alternatively, if the market does not take into account the changes in accounting method then one can see differences in company value that are unrelated to a rational estimate of changes in future cash flow. This phenomenon is often referred to in the finance literature as the functional fixation hypothesis, the naive investor hypothesis or the mechanistic investor hypothesis. Further discussions on this issue are given in Walker (2004).

The second potential reason for the weak return-earnings association is mentioned in the theoretical paper of Dye and Sridhar (2002). The authors show that there could be rational share price movements as a result of the market possessing information about the valuation implications of a firm's change in strategy that is not known by the management of the firm proposing the change. Here, managers can assess the stock market's information by announcing a potential new strategy, and then conditioning the

[18] Details on these reasons are given in Kothari (2001) and Walker (2004).

decision to implement this strategy on the size of the market's price reaction to the announcement. To date, however, no empirical evidence exists on this issue.

The third potential reason for the weak return-earnings association is the existence of value-irrelevant noise in reported earnings. The most crucial reason for this noise is the correct application of GAAP regimes. For example, many GAAP regimes allow recognising some expenses earlier while they allow recognising earnings later. GAAP regimes require companies to treat R&D as revenue expenditure. As a result, this treatment will tend to bias reported earnings downwards relative to permanent earnings specifically for growing firms with increasing levels of expenditure of R&D. Also, many GAAP regimes cause a delay in earnings recognition. They do not allow the recognition of revenue until there is a compulsory contract between the firm and its customer. In addition, GAAP regimes require companies to report certain types of gain and losses, such as a loss on the disposal of a subsidiary, at once in income statement at the end of the financial year. This mean reported earnings will include large transitory items in a specific year.

The final reason for the weak return-earnings association is earnings' lack of timeliness. The present study focuses on lack of timeliness in reported earnings and examines the extent to which this lack of timeliness is associated with forward-looking disclosures in annual report narratives.

The literature on earnings' timeliness is now reviewed. Beaver, Lambert and Morse (1980) present the first systematic evidence that US stock prices contain a richer information set than the past time-series of earnings. The importance of their study is partly derived from the introduction of a formal model of the price-earnings relation. Theoretically, they show that a price-based forecasting model yields a more accurate

prediction of future earnings changes than a model based solely on the past time-series of earnings. In the empirical part of Beaver et al. (1980), they invert the traditional price-earnings relation and test for the information content of prices with respect to future earnings. Their empirical findings indeed show a significant coefficient on the lagged price variable, consistent with the notion of prices leading earnings.

Stock prices anticipate future earnings changes because events are reflected in stock prices as soon as the information reaches the market while their influence on earnings often occurs with a time lag. This lagged response of earnings is a result of certain accounting principles such as reliability, objectivity and conservatism. These principals prevent earnings from reflecting value-relevant information in a timely manner. For example, with its emphasis on historical-cost accounting measurement, the expected future net cash flows from a new investment are not reflected in earnings in the current period. However, information about the investment causes immediate revisions in the market's expectations of future earnings and, hence, price changes.

Following Beaver Lambert and Morse (1980), a number of researchers have provided further evidence that stock prices anticipate future earnings changes. A list of these papers could include Beaver, Lambert and Ryan (1987), Freeman (1987), Collins and Kothari (1989), Kothari (1992), Kothari and Sloan (1992), Warfield and Wild (1992), Jacobson and Aaker (1993), Collins et al. (1994) and Donnelly and Walker (1995).

Beaver, Lambert and Ryan (1987) build upon the empirical analysis in Beaver, Lambert and Morse (1980) and provide results broadly consistent with those of Beaver, Lambert and Morse (1980). They find that stock returns in one year contain information about earnings in the following years. Furthermore, their regression results indicate that price-based models outperform time-series models for large firms but not for small firms.

Freeman (1987) provides empirical evidence that the information incorporated in stock prices systematically differs between large and small firms. The regression results of Freeman (1987) show that the stock prices of large firms anticipate future earnings earlier than the stock prices of small firms.

Collins and Kothari (1989) examine the intertemporal and cross-sectional determinants of Earnings Response Coefficients (ERCs). They use reverse regressions in the analysis to reduce the errors-in-variables problem. They hypothesise and document four factors contributing to cross-sectional and intertemporal differences in the ERC. They find that the ERC is positively associated with earnings persistence and economic growth opportunities. They also show that the ERC is negatively associated with the risk-free interest rate and CAPM beta risk. Also, they find that the ERC is negatively related to the interest rate through time.

Collins and Kothari (1989) also show that the earnings-return association is affected by firm size, where firm size is used as a proxy for information environment differences among large and small firms. If there are differences in the information environment, they will affect the extent to which a change in stock prices will anticipate a change in future earnings. Moreover, they argue that the conventional regression that models returns over the 12-months period seriously underestimates the extent of the association between returns and earnings news. They also demonstrate an improvement in this relation, by starting the return measurement period earlier than the contemporaneous fiscal period.

Kothari (1992) investigates the relation between price-earnings in levels and changes. The analysis in Kothari's (1992) paper shows that if one assumes that prices do not lead earnings, then the degree of bias in the ERC and the explanatory power of the price-

earnings models specified in level or change would rank identically. However, if it is assumed that prices lead earnings, then the level specification yields a less biased ERC and higher explanatory power than the change specification. Kothari (1992), therefore, suggests that the level of earnings might be the best available variable in the price-earnings regression. This is true because an accurate proxy for the market's expectation is difficult to obtain.

In an attempt to test the ideas in Kothari (1992) empirically, Kothari and Sloan (1992) assume that stock returns over a period reflect the market's revision of expected future earnings. Accounting earnings over the same period cannot reflect such revised expectations. Accordingly, the authors suggest that returns are expected to lead earnings changes. They find that US stock prices anticipate future earnings changes up to four years ahead.

Donnelly and Walker (1995) investigate the extent to which share prices anticipate future earnings changes by estimating ERCs on a sample of 179 UK companies during the period 1972-1990. Their regression results indicate that UK stock prices anticipate future earnings changes three years ahead. In other words, they find that the extent to which prices anticipate earnings in the UK is less than that reported by Kothari and Sloan (1992) for US companies. The reason for this difference may be due to differences in the informational environment or due to differences in GAAP between the UK and US.

In a further attempt to improve the weak return-earnings association, Warfield and Wild (1992) and Jacobson and Aaker (1993) include the next period's earnings as an

additional explanatory variable in their regression model.[19] Their results show that the R^2s without the next period's earnings are 0.39%, 2.44%, 5.41% and 21.21% for quarterly, semi-annual, annual and biannual reporting periods, respectively. These R^2 values increase to 1.26%, 4.41%, 15.71% and 29.28% when including the next period's earnings in the return-earnings regression model. However, Collins et al. (1994) argue that including future periods' earnings in the return-earnings regression model is subject to an errors-in-variables problem that biases the ERCs and explanatory power downward. This kind of problem is discussed in detail below.

Collins, Kothari, Shanken and Sloan (1994)

The following paragraphs outline the return-earnings regression model that is used to test the main hypotheses in the present study. The effect of disclosure quality on share price anticipation of earnings is investigated by looking at future ERCs. Future ERCs are also used in Gelb and Zarowin (2002) and Lundholm and Myers (2002).

The article by Collins et al. (1994) is a response to Lev (1989), who notes that the association between returns and current earnings is relatively weak. They investigate two potential factors contributing to the low contemporaneous return-earnings association. These factors are: (1) earnings' lack of timeliness in capturing value-relevant events (e.g., Kothari, 1992; Kothari and Sloan, 1992) and (2) the presence of irrelevant noise in reported earnings (e.g., Beaver et al., 1980). To capture the intuition that prices lead earnings, they expand the simple return-earnings regression to include future earnings growth variables. Next paragraphs are provided on the relevant ideas in Collins et al. (1994).

[19] Other attempts are made to improve the weak return-earnings association by incorporating analyst forecast information into the return-earnings model (e.g., Liu and Thomas, 2000) or by including management forecasts of earnings in the return-earnings regression model (e.g., Ota, 2001). These studies are reviewed in Ota (2003).

Collins et al. (1994: 295) motivate their multiple regression model by assuming the following return-generating process:

$$R_t = \beta_0 + \beta_1 UX_1 + \sum_{k=1}^{N} \beta_{k+1} \Delta E_t(X_{t+k}) + e_t \quad (1)$$

where R_t is the stock return for period *t*, X_t is the growth rate of earnings in period t, $UX_t = X_t - E_{t-1}(X_t)$ is the unanticipated earnings growth rate and ΔE_t is the revision in market expectations between the beginning and the end of period *t*. k is limited to three years ahead. This is motivated by Kothari and Sloan (1992), who show that share price anticipation of earnings is not significant beyond three years.

Collins et al. (1994) suggest that returns in period *t* are generated by three components: (1) the unanticipated component of the current period's earnings change, UX_t, (2) the market's revision in expectations about future earnings growth rates, $\Delta E_t(X_{t+k})$ and (3) an orthogonal error term that captures all other influences.

To implement equation (1) empirically, one needs to replace unobservable expectations with observable proxy variables. Prior to Collins et al. (1994), researchers such as Warfield and Wild (1992) use realized earnings growth as an observable proxy for the market's expectations to explain stock returns. Equation (2) shows the Warfield and Wild's regression model.

$$R_t = b_0 + b_1 X_t + \sum_{k=1}^{N} b_{k+1} X_{t+k} + e_t \quad (2)$$

Collins et al. (1994) point out that the use of realised earnings growth rates introduces errors-in-variables problems that bias the slope coefficients and R^2 downward. The

errors-in-variables problems become apparent when one rewrites Equation (2) in terms of variables of interest and measurement errors (Collins et al., 1994: 296):

$$R_t = b_0 + b_1[UX_t + E_{t-1}(X_t)] + \sum_{k=1}^{N} b_{k+1}[\Delta E_t(X_{t+k}) + UX_{t+k} + E_{t-1}(X_{t+k})] + e_t \quad (3)$$

where UX_t is the unanticipated component of current earnings growth, $E_{t-1}(X_t)$ is the portion of current period's earnings growth that is anticipated in period *t*–1, $E_{t-1}(X_{t+k})$ is the portion of period *t*+*k*'s earnings growth that is anticipated in period *t*–1, UX_{t+k} is the component of period *t*+*k*'s earnings growth generated by surprises in periods *t*+1 to *t*+*k*. Comparing equation (2) with equation (3), it can be seen that equation (2) gives rise to a number of measurement error problems. Firstly, X_t differs from UX_t by the expectations from $E_{t-1}(X_t)$. Secondly, X_{t+k} differs from UX_{t+k} in two aspects. First, the market may already know information about X_{t+k} at time point *t*–1. In other words, the parameter associated with $E_{t-1}(X_{t+k})$ may be non-zero. Second, new information about X_{t+k} may be available to the market between time point *t* and time point *t*+1. This is indicated by the term UX_{t+k}.

An important observation in Collins et al. (1994) is that one can mitigate these measurement error problems by the inclusion of errors-in-variables proxies in the augmented regression model. Crucially, Collins et al. (1994) establish that the inclusion of such proxies will affect the goodness of fit of the model, only if the reason for the poor performance of the simple return-earnings regression is 'prices leading earnings'. If value-irrelevant noise is the cause of the poor statistical performance of the standard return-earnings model, then the goodness of fit of Equation (2) will not be improved by adding these proxies.

Collins et al. (1994) suggest three measurement error proxies. These are lagged earnings yield, EP_{t-1}, current growth in book value of assets, AG_t and future periods' returns, R_{t+k}. Including these proxies in equation (2) yields the following expanded regression model[20]:

$$R_t = b_0 + b_1 X_t + \sum_{k=1}^{N} b_{k+1} X_{t+k} + \sum_{k=1}^{N} b_{k+N+1} R_{t+k} + b_{2N+2} EP_{t-1} + b_{2N+3} AG_t \quad (4)$$

The economic rationale for using the three proxies is summarized in the following paragraphs. These arguments follow Collins et al. (1994).

The first measurement error proxy for expected future earnings growth is the lagged earnings yield variable, EP_{t-1}. This variable is defined as period *t*–1's earnings over price at the start of the return window for period *t*. Given that price impounds information about future earnings, EP_{t-1} proxies for the market's forecast of further earnings growth [i.e., proxies for $E_{t-1}(X_t)$ and $E_{t-1}(X_{t+k})$]. It is well known that prices incorporate information about future earnings. Therefore, a high price in relation to last year's earnings signals high expected earnings growth for the current and future years. As the earnings yield variable and expected earnings growth (the measurement error) are negatively associated, the coefficient on EP_{t-1} should be positive. This is true because this proxy serves to subtract the noise element from realised earnings growth.

The second proxy is the asset growth variable, AG_t. Higher asset growth indicates that managers increase their production capacity due to an expectation of a higher demand for their product in the future. Such an expansion should lead to higher expected

[20] Equation (4) corresponds to Collins et al.'s (1994:297) Equation (6).

earnings growth. Given that asset growth and expected future earnings changes are positively associated, the coefficient on AG_t is forecasted to be negative.

Finally, the measurement error proxy for UX_{t+k} is future periods' returns, R_{t+k}. Unanticipated future events that lead to higher (lower) earnings growth in period *t+k* should also lead to positive (negative) returns in the period when the news becomes available to the market. Hence, a positive relation between UX_{t+k} and future returns is expected to result in negative coefficients on the return variables in Equation (4).

Collins et al.'s (1994) empirical tests of the noise versus lack of timeliness hypotheses use return-earnings data at three levels of aggregation: economy level, industry level and firm level. This is due to the fact that the implied effect of data aggregation on the contemporaneous return-earnings relation is different under the two hypotheses. The authors argue that, as data are aggregated the contemporaneous association between returns and earnings should strengthen under the noise-in-earnings hypothesis, but not under the earnings' timeliness hypothesis. As data are aggregated, the weakly cross-sectionally correlated noise will be diversified away. Consequently, only the value-relevant earnings component in the aggregated data will remain and this would be highly correlated with contemporaneous stock returns. On the other hand, under the lack-of-timeliness hypothesis, cross-sectional data aggregation would not be helpful in making current earnings more timely with respect to contemporaneous returns. This means that the contemporaneous return-earnings association is not expected to improve upon aggregation under the lack-of-timeliness hypothesis.

Collins et al.'s (1994) modifications to the traditional contemporaneous return-earnings regression model yield large increases in the explanatory power. In particular, the R^2 of the contemporaneous return-earnings model was generally under 15%. The inclusion of

future earnings growth variables in the return-earnings regression increases the R^2 to approximately 35-50%.

Collins et al. (1994) show that earnings' lack of timeliness is the most important factor contributing to the low contemporaneous return-earnings relation. Since current and future earnings measures are useful in explaining current returns, Collins et al.'s (1994) results confirm that accrual-based earnings measures do not capture value-relevant events or market expectations in a timely manner. On the other hand, Collins et al. (1994) find that noise in earnings does not appear to be a major factor in explaining the low contemporaneous return-earnings association.

3.3. Economic Consequences of Corporate Disclosure

This section reviews the available evidence on the economic consequences of corporate disclosure. The literature has examined the effects of corporate disclosure on the cost of capital, analysts' forecasts, analyst following, stock price liquidity, bid-ask spreads and earnings timeliness. Table 3.1 provides a summary of these studies in chronological order.

The main purpose of introducing Table 3.1 is to demonstrate the marked differences in sample sizes and disclosure proxies that have been used in prior research in the UK compared to non- UK studies. Therefore, the main focus of the discussion that follows is the information which appears in Columns 3, 4 and 5.

Table 3.1. Summary of Research on the Economic Consequences of Corporate Disclosure

Reference	Research Issue	Sample Size	Disclosure Proxy	Country	Findings
Lang and Lundholm (1993)	Determinants of analyst ratings of corporate disclosures.	2,319 firm-years	Subjective ratings based on AIMR-FAF ratings	USA	Disclosure scores are higher for large firms with a weaker relation between stock returns and earnings.
Welker (1995)	Association between corporate disclosure quality and the cost of debt.	1,639 firm-years	Subjective ratings based on AIMR-FAF ratings	USA	Negative association between disclosure scores and cost of debt.
Lang and Lundholm (1996)	The association between disclosure quality and analyst following and the properties of analyst forecasts.	751 firms	Subjective ratings based on AIMR-FAF ratings	USA	High disclosure firms have a large analyst following, more accurate analyst earnings forecasts, less dispersion between analysts and less volatility in forecast revisions.
Schleicher (1996)	Effect of the quality of UK annual report disclosures on share price anticipation of earnings.	18 firms (200 firm-years)	Self-constructed index	UK	No association between the quality of corporate annual reports and share price anticipation of earnings.
Botosan (1997)	Association between disclosure quality and the cost of equity capital.	122 firms	Self-constructed index	USA	Negative association between disclosure scores and cost of capital.
Bryan (1997)	Information content of the MD&A.	250 firms	Self-constructed index	USA	Forward-looking disclosures about operations in the MD&A are significantly associated with one-period ahead change in earnings.

Francis et al. (1997)	Effect of management communications with securities analysts on analyst behaviour.	200 firm-presentations	Self-constructed index based on presentations available on *Investext Plus*	USA	Management communications with analysts lead to an increase in analyst following, a positive abnormal return on the presentation date. No evidence that such presentations increase the analyst forecasts accuracy.
Sengupta (1998)	Association between corporate disclosure quality and the cost of debt.	725 firm-years	Subjective ratings based on AIMR-FAF ratings	USA	Increased disclosure leads to a lower cost of debt.
Barron et al. (1999)	Association between the quality of the MD&A and analysts' earnings forecasts.	284 firms	Self-constructed index based on SEC ratings	USA	Firms with high MD&A ratings experience less error and less dispersion in analyst forecasts. This finding is driven by forward-looking disclosures about capital expenditure and operations and historical disclosure about capital expenditure.
Brown et al. (1999)	Effect of disclosure on analyst forecasts and share price anticipation of earnings.	727 firms	Subjective ratings based on Australian Stock Exchange	Australia	Analyst forecasts are not affected by the introduction of statutory sanctions. Furthermore, after such introduction stock prices anticipate future earnings change earlier only for small firms.

Eng and Teo (1999)	Effect of corporate disclosure on analyst behaviour.	125 firms	Self-constructed index	Singapore	Firms with high disclosure scores experience an increase in analyst forecasts' accuracy, an increase in the number of analyst following and a decrease in dispersion in analysts' earnings forecasts.
Frankel et al. (1999)	Benefits of corporate conference calls.	1056 corporate conference calls	Information in conference calls.	USA	Firms holding conference calls tend to be followed by financial analysts. Conference calls reduce information asymmetry.
Healy et al. (1999)	Economic consequences of increased corporate disclosure.	595 firms	Subjective ratings based on AIMR-FAF ratings	USA	Firms with increasing disclosure levels over time experience improved stock performance and capital market intermediation.
Schleicher and Walker (1999)	Effect of voluntary disclosure, on the informativeness of stock prices.	(20 firms) 220 firm-years.	Self-constructed index	UK	Stock prices are better informed when annual reports capture future oriented information.
Botosan and Harris (2000)	Determinants of managers' decisions to increase segment disclosure frequency	107 firms	Quarterly segment disclosure	USA	Firms that change their frequency of segment reporting experience an increase in analyst following.
Bushee and Noe (2000)	Association between disclosure scores and return volatility	4,314 firm-years	Subjective ratings based on AIMR-FAF ratings.	USA	Higher disclosure firms have greater institutional ownership. No impact on return volatility.

Doula (2000)	Effect of forward-looking information in annual report on price leading earnings.	48 firms	Self-constructed index:	UK	Some evidence that disclosing forward-looking information such as sales, profits, capital and acquisition leads to more informative stock prices.
Kanto and Schadewitz (2000)	Effect of individual classes of information in interim reports on the return-earnings relation.	573 interim reports	Self-constructed index	Finland	Information contained in interim reports is value-relevant over and above that contained in earnings.
Leuz and Verrecchia (2000)	Economic consequences of increased disclosure.	102 firms	Previous literature	Germany	High level of disclosure is associated with lower bid-ask spread. Higher level of disclosure has no effect on share price volatility.
Rashid (2000)	Effect of disclosure quality on cost of debt.	41 retail banks	Self-constructed index	UK	Disclosure quality is negatively associated with the cost of debt.
Barron et al. (2001)	Effect of MD&A disclosures on stock prices	84 firms	Subjective ratings based on SEC ratings	Canada	MD&A disclosure contains incremental price relevant information.
Brown et al. (2001)	Relation between disclosure quality and information asymmetry.	275 firms	Subjective ratings based on AIMR-FAF ratings.	USA	Disclosure is negatively associated with the level of information asymmetry.
Hefli et al. (2001)	Relation between disclosure quality and market liquidity.	298 firms	Subjective ratings based on AIMR-FAF ratings	USA	Higher disclosure firms have lower effective bid-ask spread and lower adverse selection spread. Higher disclosure enhances a firm's market liquidity.

Richardson and Welker (2001)	Relation between disclosure quality and the cost of capital.	124 firms	Subjective ratings based on SMAC ratings	Canada	Disclosure ratings are negatively related with cost of capital for firms with low analyst following.
Walker and Tsalta (2001)	Relation between corporate voluntary disclosure and analyst following.	57 firms	Self-constructed index	UK	Positive association between the quality of forward-looking information in annual reports and analyst following.
Bowen et al. (2002)	Effect of conference calls on analysts' forecasts.	12,555 firm-quarters	Conference calls	USA	Conference calls increase analyst forecast accuracy and agreement between analysts. It is also useful for analysts with weak prior forecasting accuracy.
Blasco and Trombetta (2002)	Relation between the quality of corporate disclosure and a firm's reputation.	155 firm-years	Subjective ratings based on 'Actualidad Economica'	Spain	Quality of annual report disclosure increases a firm's reputation after controlling for size.
Botosan and Plumlee (2002)	Effect of disclosure level on cost of equity capital.	3,620 firm-years	Subjective ratings based on AIMR-FAF ratings	USA	Negative relation between annual report disclosure quality and cost of capital. Positive association between quarterly disclosure quality and cost of capital. No association between investor relation quality and cost of capital.
Gelb and Zarowin (2002)	Effect of disclosure policy on the informativeness of stock prices for future earnings.	821 firms	Subjective ratings based on AIMR-FAF ratings	USA	High level of disclosure is associated with stock prices that are more informative about future earnings changes.

Hail (2002)	Effect of voluntary disclosure on cost of capital	37 non-financial firms	Subjective ratings based on 1- Swiss Banking. 2- Financial analysts' opinion. 3- Investors needs.	Switzer-land	Negative association between disclosure and cost of capital.
Haw et al. (2002)	Effect of corporate disclosure on prices leading earnings.	1,349 firm-years	Subjective ratings based on AIMR-FAF ratings	USA	One year-ahead share price anticipation of earnings increases with the informativeness of investor relations but does not vary systematically with the informativeness of annual and quarterly reports.
Lundholm and Myers (2002)	Effect of corporate disclosure policy on the informativeness of stock prices.	724 firms	Subjective ratings based on AIMR-FAF ratings	USA.	High level of disclosure is associated with stock prices that are more informative about future earnings changes. Changes in disclosure levels are positively related to changes in the importance of future earnings news for current returns.
Schadewitz and Kanto (2002)	Effect of interim report disclosures on the return-earnings relation.	573 interim reports	Self-constructed index	Finland	For high disclosure firms market responses to interim reports is quicker than for medium and low disclosure firms.
Bulter et al. (2003)	Effect of the frequency of financial reports on the timeliness of earnings.	3,702 firms	Firm's actual reporting frequency less its SEC-required reporting frequency	USA	Stock prices anticipate future earnings changes quicker for firms with quarterly reporting than for firms with semi-annually reporting.

Byard and Shaw (2003)	Effect of disclosure quality on analyst forecasts.	1,017 firms	Subjective ratings based on AIMR-FAF ratings	USA	High level of disclosure is positively associated with analyst forecasts accuracy. No association between the quality of private communications with analysts and analyst forecasts accuracy.
Hope (2003a)	Relation between disclosure practice and analyst forecast accuracy.	896 firms	Subjective ratings based on CIFAR ratings	Different countries	Firms' level of disclosure is positively related to forecast accuracy.
Hope (2003b)	Association between accounting policy disclosure and analysts' forecasts.	811 firms for examining forecast error. 783 for examining forecast dispersion	Subjective ratings based on CIFAR ratings	Different countries.	Level of accounting policy disclosure is negatively related to forecast dispersion and forecast error.
Jr. et al. (2003)	Relation between different levels of disclosures by foreign companies listed on US exchanges and the earnings-return association.	536 firm-years	Item 17 and Item 18 disclosure rules under US GAAP	USA	Higher levels of disclosure are associated with higher earnings response coefficient (ERCs).
Kothari and Short (2003)	Effect of disclosures by management, analysts and financial press on the cost of capital.	887 firms	Self-constructed index	USA	Positive business press news decreases the cost of capital, while negative news disclosure increases it.

Shaw (2003)	Association between disclosure quality and the return-earnings association.	1,113 firm-years	Subjective ratings based on AIMR-FAF ratings	USA	For bad news years the contemporaneous return-earnings relation is inversely related to disclosure quality. There is no association between the return-earnings correlation and disclosure quality in good news years.
Shuqing et al. (2003)	Effect of disclosure quality, ownership structure and propriety cost on the return-earnings association.	172 firms (516 firm-years)	Self-constructed index	Singapore	High level of disclosure is associated with stock prices that are more informative about future earnings changes. However, this association is weaker if management holds a higher proportion of share ownership, if government ownership exists, and/or if proprietary costs exist.

3.3.1. Disclosure Quality and Cost of Capital

Two theoretical frameworks support the proposition that greater disclosure is associated with a lower cost of capital. The first suggests that greater disclosure reduces the possibility of information asymmetries between the firm and its shareholders or among potential buyers and sellers of a firm's shares. This, in turn, should reduce the cost of capital by reducing the discount at which shares are sold through reduced transaction costs or by increasing the demand for a firm's securities (e.g., Amihud and Mendelson, 1986; Diamond and Verrecchia, 1991). The second framework suggests that greater disclosure reduces the estimation risk associated with investors' assessments of the

parameters of the return or payoff distribution and, thereby, reduces the cost of capital (e.g., Coles et al., 1995 and Clarkson et al., 1996).

A number of recent US studies support this theory by examining the link between disclosure and a firm's cost of capital. For example, Welker (1995) and Sengupta (1998) document that firms with higher disclosure ratings have, on average, lower bid-ask spreads and lower cost of debt at the time of issue, respectively. Botosan (1997) shows a negative association between disclosure levels and the firm's cost of equity capital, but only for firms with a low analyst following. Healy et al. (1999) explore the effect of a sustained increase in disclosure on a number of variables expected to be associated with the cost of equity capital. For example, they show that firms that increase their disclosure levels experience improvements in the bid-ask spread.

Botosan and Plumlee (2002) re-examine the relation between corporate disclosure and the cost of equity capital. Contrary to their expectations, Botosan and Plumlee (2002) find that greater *total* disclosure is associated with a higher cost of equity capital. Therefore, they examine the extent to which the relationship between disclosure and the cost of capital varies by type of disclosure. They find a negative association between annual report disclosures and the cost of capital and a positive association between quarterly report disclosures and the cost of capital. They do not find any association between investor relations and the cost of capital. Hail (2002) also explores the relation between disclosure quality and the cost of equity capital for Swiss companies. He shows a strong negative association between disclosure and cost of capital. Unlike Botosan (1997) Hail's results are not restricted to low analyst following only.

Finally, Rashid (2000) investigates the link between disclosure and the firm's cost of debt equity for retail banks listed on the London Stock Exchange. He finds that firms with higher disclosure scores experience lower cost of debt capital.

Overall, the above studies show an inverse relation between disclosure and cost of capital. The current study examines the benefits of corporate disclosure from a different angle. It tests whether increased levels of voluntary disclosure in corporate annual reports allow a better informed stock market. The concept of a 'better informed stock market' refers to the market's ability to anticipate future earnings changes more accurately.

3.3.2. Disclosure Quality and Analyst Behaviour

Another economic consequence of corporate disclosure that has received much attention refers to analyst behaviour. Empirical support for the association between disclosure quality and analyst behaviour comes from a number of studies.

Lang and Lundholm (1996) is the first study that examines the association between disclosure and analyst behaviour. They start by looking at analyst following which is defined as the number of analysts following each firm. Lang and Lundholm argue that increased corporate disclosure is expected to increase the supply of analyst services. However, the demand for analyst services will depend on the role that analysts play in the capital market. If analysts act mainly as information intermediaries who process firm provided information for ordinary investors, then an increase in firm provided information will ensure that the analyst has a more valuable report to sell. In this case increased disclosure leads to an increase in the demand for analysts' services. However, if analysts act mainly as information providers competing with firm-provided disclosures made directly to investors, then an increase in the level of disclosure will

substitute for the analysts' reports. In this case, increased disclosure leads to a decrease in the demand for analysts' services.

The second aspect of analyst behaviour in Lang and Lundholm (1996) refers to analyst earnings forecasts which are measured by forecast accuracy, the degree of dispersion among individual analyst forecasts and the variability of forecasts revisions during the year. Lang and Lundholm (1996:471) argue that the effect of increased disclosure on the dispersion of analyst forecasts depends on whether differences in forecasts are due to differences in information or differences in forecasting models. They state that if analysts have a common forecasting model and observe the same firm-provided disclosures but process different private information, they will place less weight on their private information as the informativeness of firm provided disclosure increases, and thus the consensus among their forecasts will increase. If analysts have the same firm-provided and private information but differ in the weights they place on components of firm-provided disclosure in forecasting earnings, then additional disclosure might increase the dispersion of analyst forecasts. The likely relation between corporate disclosure and analyst forecast accuracy is clearer. They predict that analyst forecast accuracy would increase with the quality of a firm's disclosure policy. Finally, they expect that the volatility of forecast revisions in the period up to an earnings announcement is likely to be reduced by increasing the level of corporate disclosure.

Lang and Lundholm's (1996) main results indicate that there is a significant positive association between corporate disclosure ratings and analyst following. They also show that increases in disclosure quality tend to be followed by higher levels of analyst following. They do not find evidence that increased analyst following leads to an increase in corporate disclosure. Their results also show that firms with more

forthcoming disclosures have more accurate consensus forecasts, less dispersion among individual analyst forecasts and less variable forecast revisions.

Following Lang and Lundholm (1996) a number of other studies have investigated the association between disclosure quality and analyst behaviour. For example, Byard and Shaw (2003) find in the US that the level of corporate disclosure quality is positively associated with analyst forecast accuracy. Walker and Tsalta (2001) find a strong positive relation between the quality of forward-looking information in the UK annual report discussion section and analyst following. Eng and Teo (2000) show that greater corporate disclosure by Singapore firms lead to greater analyst following, more accurate analyst forecasts and less dispersion in the earnings forecasts among analysts.

Other groups of studies focus on the effect of mandatory disclosure on analyst earnings forecast accuracy. Some of these studies find that increases in the level of required disclosure leads to an increase in analyst forecast accuracy (e.g., Brown and Han, 1992; Swaminthan, 1991). On the other hand, Brown et al. (1999) find that the introduction of new accounting regulations does not affect analyst forecasts.

Researchers have also investigated the effect of different channels of disclosure on analyst forecasts accuracy. For example, Francis et al. (1997) and Byard and Shaw (2003) show that management's direct communications with financial analysts have no effect on analysts' forecast accuracy or dispersion. Bowen et al. (2002) show that conference calls increase analysts' ability to forecast earnings accurately and that conference calls decrease dispersion among analysts. Also, analysts with relatively weak forecasting performance benefit more from conference calls.

Finally, by using quantitative management earnings forecasts as a proxy for disclosure quality, a number of studies highlight the importance of corporate disclosure to financial

analysts. The results of these studies show that financial analysts revise their forecasts after management earnings forecasts are released (see Baginski and Hassell, 1990; Jennings, 1987; Williams, 1996; Waymire, 1986; and Kross et al, 1990).

In summary, existing empirical evidence indicates that enhanced disclosure is associated with more accurate analysts' forecasts. Although such results are important, the current study seeks to test the relation between prices and future earnings directly rather than relying on proxies such as analyst forecasts. This is because more accurate analyst forecasts might be evidence of firms 'managing' their analyst relationships better rather than evidence of more informative prices (Gelb and Zarowin, 2002). The following section reviews the literature related to the association between corporate disclosure and earnings timeliness.

3.3.3. Disclosure Quality and Timeliness of Earnings

The relation between corporate disclosure quality and the return-earnings relation has been a topic of considerable interest among accounting researchers. Schleicher (1996) find that there is no association between the quality of annual report disclosure and the informativeness of stock prices about future earnings changes. Schleicher and Walker (1999) find that the inclusion of forward-looking information in the OFR statement is useful in anticipating future earnings changes. Using AIMR-FAF ratings as a proxy for disclosure quality Healy et al. (1999) find that high disclosure firms experience an increase in the coefficient on current earnings in a regression of current returns on current earnings. However, without future earnings in the regression, they argue that the coefficient on current earnings must capture changing expectations about future earnings.

More recently, Lundholm and Myers (2002), Gelb and Zarowin (2002) and Haw et al. (2002) find that increased disclosure is associated with stock prices that are more informative about future earnings change.

Finally, Shaw (2003) estimates reverse regressions of earnings on returns and interacts the return variable with disclosure quality scores and news type. The main finding is that disclosure quality has little impact on the earnings-return relation in good news years.

Below, the key papers that are directly relevant to the empirical analysis of the present study are reviewed. These studies are Schleicher (1996), Schleicher and Walker (1999), Lundholm and Myers (2002), Gelb and Zarowin (2002) and Haw et al. (2002).

Schleicher (1996)

Schleicher (1996) is the first study to examine the effect of corporate annual report disclosures on prices leading earnings. To achieve this he constructs a comprehensive disclosure index. The selection of items included in his index is guided by recommendations provided in the ASB statement 'Operating and Financial Review', the Companies Acts, the Yellow Book, The Corporate Report (ASSC, 1975), 'Making Corporate Reports Valuable (ICAS, 1988) and previous disclosure studies. The disclosure index in Schleicher (1996) contains 404 topics.

One of the aims of Schleicher's (1996) study is to assess the impact of the quality of annual report disclosure on the informativeness of stock prices. The author hypothesises that the extent to which prices anticipate future earnings changes is positively associated with the quality of corporate annual reports.

Schleicher's (1996) study is limited to a small number of firms in comparison with US studies. This is due to the time-consuming nature of the work involved in assessing annual reports disclosures via manual reading. The results in Schleicher's (1996) study are based on a sample of eighteen non-financial UK firms represented by 162 firm-year observations. The author uses the theoretical model introduced by Kothari (1992) and Donnelly and Walker (1995) to test his hypothesis.

Schleicher's (1996) study does not find a significant association between the level of annual report disclosure scores and the informativeness of stock prices. One possible explanation is that his all-inclusive disclosure index contains some topics which are not value-relevant to investors. Such topics would not increase the market's ability to anticipate future earnings changes. Therefore, these topics introduce noise into the measure of disclosure quality.

Schleicher and Walker (1999)

The work in Schleicher and Walker (1999) is an extension of the work in Schleicher (1996). To measure the quality of the annual report discussion section, Schleicher and Walker (1999) construct their disclosure indices based on the methodology in Schleicher (1996). The final outcomes are three equally weighted indices. These indices combine 82 items in the Operating and Financial Review (DOFR) index, 64 items in the Operating and Financial Projections (DOPF) index, and 34 items in the Segmental Reporting (DSEG) index.

The Schleicher and Walker study is limited to a small number of firms. Their sample contained 20 non-financial UK firms (220 company-years). These firms are classified into three industry sectors (engineering, electronic and electrical equipment).

The regression results in Schleicher and Walker (1999) show that higher levels of narrative disclosures in the annual report enable the market to better anticipate future earnings changes. This effect is particulary strong when they use the forward-looking disclosure index, DOPF, and when they use models that examine one-period- ahead and two-period-ahead share price anticipation.

Schleicher and Walker (1999) is the first published paper that examines the effect of disclosure quality on prices leading earnings. It provides evidence that forward-looking information in the annual report increases the market's ability to anticipate future earnings changes. However, the Schleicher and Walker (1999) study suffers from at least two major limitations. First, the sample is relatively small in comparison with US studies. Secondly, they cover only three related industry-sectors. Therefore, it is not possible to generalise their results to other sectors.

The work in Schleicher and Walker (1999) suggests that the market better anticipates future earnings changes when annual reports include forward-looking information. This finding provides a motivation for the present study to focus on forward-looking disclosures in annual report narratives. Furthermore, due to the availability of electronic annual reports, it is possible to use computerised qualitative software to automate the scoring process. Therefore, one can conduct a study with a much larger number of firms.

Lundholm and Myers (2002)

Lundholm and Myers (2002) investigate the effect of corporate disclosure on the return-earnings association. They use AIMR-FAF ratings as a proxy for disclosure quality. The sample in Lundholm and Myers (2002) consists of 724 firms in 33 industries from the period 1980-1994. At total of 4,478 firm-years are used.

Lundholm and Myers (2002) use the regression model from Collins et al. (1994). This model regresses current stock returns on both current and future earnings changes. As discussed earlier, a main contribution of Collins et al. (1994) is the inclusion of future returns in the augmented return-earnings regression to control for the unexpected component of future realised earnings.

Lundholm and Myers (2002) hypothesise that corporate disclosure is a significant source of changing expectations about a company's future performance. They argue that if firms reveal news relevant for forecasting future earnings, then news about future earnings will be reflected in current stock returns. In this case the coefficient on proxies for news about future earnings will be positive in the return-earnings regression model. In contrast, if firms do not disclose news about future earnings, then such news will not be revealed to the market. In this case, the coefficient on future earnings will be close to zero. This means that there is an interaction effect between future earnings and disclosure levels. Accordingly, the authors test whether 'current returns are increasing in revealed future earnings as measured by the interaction between the level of disclosure and the realized future earnings' (p.814).

Lundholm and Myers (2002) also examine the effect of the quality of corporate disclosure on the importance of current earnings news. If increases in corporate disclosure cause current returns to depend more heavily on future earnings news, then current earnings news might become less relevant. Similarly, current earnings might appear to be value-relevant for low disclosure firms because current earnings proxy for changes in expectations about future earnings. Their second hypothesis is that 'current returns are increasing in current earnings, but at a decreasing rate as disclosure increases' (p.816).

The third hypothesis in Lundholm and Myers (2002) is related to the association between changes in corporate disclosure and changes in earnings' timeliness. If disclosure quality affects the extent to which future earnings are incorporated in current returns, then the amount of variation in current returns that is due to future earnings should be increasing in the level of disclosure quality. This argument leads to the hypothesis that 'increases (decreases) in firms' disclosure activity lead to more (less) informative stock prices' (p. 817).

The results show that the disclosure levels are positively associated with the market's ability to predict future earnings changes. However, the importance of current earnings for stock returns is not related to disclosure quality. In relation to the time-series analysis, the regression results indicate that increases in disclosure lead to higher levels of share price anticipation of earnings, as predicted.

Gelb and Zarowin (2002)

Gelb and Zarowin (2002) also examine the relation between the disclosure practices of firms and the phenomenon of share price anticipation of earnings. Like Lundholm and Myers (2002) they use the AIMR-FAF corporate disclosure ratings as a proxy for disclosure quality.

Gelb and Zarowin (2002) collect their data on disclosure quality from the 1980-1993 AIMR-FAF reports. The total number of firms included in their study is 821 non-financial firms which are classified into 22 separate industry sectors. The authors then divide their sample into two categories. The two categories include firms with high and low AIMR-FAF disclosure scores as defined in terms of top versus bottom quartile in their industry for two consecutive years.

The authors hypothesise that high disclosure firms experience greater price informativeness than low disclosure firms. They define price informativeness by the association between current stock returns and future earnings changes.

Like Lundholm and Myers (2002), Gelb and Zarowin use the regression model introduced by Collins et al. (1994). They regress current stock returns on both current and future earnings changes. The empirical results in Gelb and Zarowin (2002) indicate that disclosure quality is positively associated with the market's ability to anticipate future earnings changes.

Gelb and Zarowin (2002) also investigate whether certain types of corporate disclosures channels make stock prices more informative for future earnings changes. They redefine their disclosure scores based on three communication channels. These include annual report disclosure scores, quarterly reports and other publication disclosure scores and the investor relations disclosure scores. They run their regression model again to test the effect of these three types of disclosures on prices leading earnings. The regression results indicate that increased disclosure in the annual report does not make stock prices more informative. These results are consistent with Schleicher (1996), who does not find a significant association between annual report disclosure scores and prices leading earnings. On the other hand, they find a significant association between both the quarterly and other publication and the investors' relation and the informativeness of stock prices.

Haw et al. (2002)

Like Gelb and Zarowin (2002), Haw et al. (2002) examine the effect of overall disclosure and different disclosure channels (i.e. annual reports, quarterly reports and investor relations) on share price anticipation of earnings.

The authors collect their disclosure scores from the AIMR-FAF analysts' ratings. Their sample size is 1349 firm-year observations covering the period between 1983-1991. They examine the association between the overall disclosure scores and earnings' timeliness. They also examine the effect of each disclosure channel score on prices leading earnings.

The authors use the regression model introduced by Freeman (1987). Their regression results show that the stock returns of firms with high disclosure incorporate earnings changes about one-year ahead of those of firms with less disclosure. They also find that higher levels of investor relations' disclosure increase the market's ability to anticipate future earnings changes, while higher levels of annual and quarterly reports disclosures do not.

3.4. The Present Study

It is clear from Table 3.1, Columns 3 and 5, that most of the empirical disclosure literature is based on US data. This is due to the availability of AIMR-FAF ratings for a large sample of firms. Furthermore, one can observe that the availability of subjective ratings in other countries, such as Australia (Australian Stock Exchange), Canada (SEC ratings) and Spain (Actualidad Economica), provide an opportunity to undertake large-scale disclosure studies.

In the UK, there are no subjective analyst ratings similar to those in the other countries. Therefore, the only way to perform a large-scale disclosure study is to construct a disclosure index to proxy for the quality of corporate disclosure. One of the main problems in self-constructed disclosure indices is that they involve a large amount of time and effort. As can be observed from Table 3.1, the sample sizes of UK disclosure studies relating to the economic consequences of corporate disclosure is relatively small

in comparison with the US studies (e.g., 18 firms in Schleicher, 1996; 20 firms in Schleicher and Walker, 1999; 48 firms in Doula, 2000; 41 firms in Rashid, 2000; and 57 firms in Walker and Tsalta, 2001). UK researchers have had a difficulty in scoring a large sample of annual reports and, hence, in undertaking large-scale disclosure studies. The sample size problem in UK disclosure studies is partly due to the labour-intensive work required to score a large sample of annual reports.

The present study contributes to the literature by introducing a new methodology for evaluating the quality of voluntary corporate disclosures in annual report narratives. An important feature of this methodology is that it uses a text analysis software package to speed up the process of identifying the list of relevant disclosure items. It also assists in automating the generation of disclosure scores for annual reports. This automation process makes it possible to score a large number of annual reports with less effort, cost and time. The resulting sample sizes are comparable to those employed by US researchers with access to annual AIMR-FAF disclosure ratings.[21]

Following Lundholm and Myers (2002) and Gelb and Zarowin (2002), the present study investigates whether a higher level of corporate disclosure allows the market to anticipate future earnings changes more accurately. Based on the regression model pioneered in Collins et al. (1994), it regresses current returns on current and future earnings variables and it allows the regression coefficients to vary with the firm's disclosure quality. Unlike Lundholm and Myers (2002) and Gelb and Zarowin (2002), however, my study emphasises individual classes of information. It examines the effect

[21] As mentioned earlier, the AIMR-FAF database is becoming increasingly dated because the *Financial Analysts Federation* discontinued the ratings in 1995. As a result, US researchers nowadays start using the computerised content analysis approach to examine the economic consequences of corporate disclosures. For example, Kothari and Short (2003) use the *General Inquirer* software to score different financial communication channels. They then examine the association between their measure of disclosure quality and the equity cost of capital.

of the publication of particular classes of forward-looking information to the market. These classes are profit topics, costs topics and turnover topics.

3.5. Summary

This chapter reviews the core literature on earnings' timeliness and the relation between earnings' timeliness and disclosure. The first part of this review is related to earnings' timeliness. The theoretical model introduced by Collins et al. (1994) is discussed in detail. This is because this model is used to examine the hypothesised association between disclosure quality and earnings' timeliness.

The second part of the review discusses the effects of corporate disclosure on various dimensions. These include bid-ask spreads, the cost of capital, analyst forecasts, analyst following, stock price liquidity and earnings' timeliness.

Because the focus of the present study is on the effect of corporate disclosure on prices leading earnings, five key papers that are the closest to this area of research were discussed in more detail in this chapter.

This chapter ends by discussing the extent to which the present study contributes to the existing Market Based Accounting Research literature. It also discusses the main differences between my study and those published in the US.

Chapter 4: The Selection of Disclosure Items: A New Methodology

4.1. Overview

The main purpose of this study is to examine the extent to which the quality of forward-looking disclosure in annual report narratives is positively associated with the stock market's ability to anticipate future earnings changes. To measure the quality of corporate disclosure, it is necessary to construct a list of items, which is sufficiently comprehensive to capture all types of relevant information to the users of annual reports. This chapter details the procedures that are followed in selecting the final list of disclosure items.

Measuring the quality of corporate disclosures is not an easy task. Academic researchers use different proxies for disclosure quality, including management forecasts (e.g., Miller and Piotroski, 2000), self-constructed measures (e.g., Schleicher and Walker, 1999) and subjective measures like AIMR-FAF ratings (e.g., Lundholm and Myers, 2002). Each proxy has its limitations that could reduce the power of tests when examining the benefits of enhanced corporate disclosures.

Core (2001) argues that there is a need to develop improved disclosure measures. He suggests that researchers can import techniques in natural language processing from fields like computer science, linguistics and artificial intelligence to construct their disclosure scores. In responding to Core's suggestion, this chapter presents a new methodology for selecting the final list of disclosure topics.

In the present study, the selection of disclosure topics is based on the types of information that are cited by sell-side analysts in their written reports. In particular, analysts' reports are searched to identify the types of forward-looking topics that are used in the analysts' decision-making process. These topics are then used to score a

sample of annual reports. A novel feature of this methodology is that, in contrast to previous studies, it focuses on the types of information that analysts actually refer to in their reports. Therefore, this study attempts to eliminate specific problems associated with previous studies, which used questionnaires and interviews to gather analysts' opinions about disclosure quality.[22]

This chapter consists of six sections, followed by a conclusion. Section 4.2 outlines the main issues regarding sell-side financial analysts. These include their definition, their types of reports, their sources of information and the content of their reports. Section 4.3 introduces the methodology that is adopted to select the disclosure items. Section 4.4 starts by introducing the data source for analysts' reports. This is followed by a discussion of issues related to sample selection. Section 4.5 discusses the necessary stages to construct the scoring sheet. Section 4.6 evaluates the methodology adopted to construct the scoring sheet. Section 4.7 summarises.

4.2. Sell-Side Financial Analysts

4.2.1. Overview

In the literature, there are two types of financial analysts, usually referred to as 'sell-side' and 'buy-side' analysts. The former group works for brokerage houses. They perform a variety of functions such as producing corporate earnings forecasts, publishing reports to the public on individual firms and issuing stock recommendations. The latter group works for those who manage large investment portfolios, such as pension funds and mutual funds.

[22] See Chapter 2 for detailed discussions.

The present study focuses on the reports of sell-side analysts.[23] These reports are used to identify topics that analysts actually used in forecasting earnings and in valuing firms.

4.2.2. Types of Analysts' Reports

Sell-side analysts write their reports for current and potential investors. In these reports, they make formal recommendations (buy, hold, or sell), present the underlying reasoning supporting their recommendations, use different valuation models to value companies and reflect upon the various factors used to discriminate between investments of different quality.

Financial analysts write two different reports, comprehensive and update reports.[24] On average, they issue one comprehensive report per year for each firm they follow, while they issue around 4 update reports per year to cover any new information after the publication of the comprehensive report. Comprehensive reports typically contain a full discussion of all aspects of the firm activities, its past performance and its expected performance. They are distinguishable from updates by virtue of the fact that the update reports are typically issued following new information or new market conditions. Update reports, in general, are much shorter than the comprehensive ones and usually point out adjustments to be made to the most recent comprehensive report.

4.2.3. Analysts' Sources of Information

The sources of information used by analysts have been investigated in a large number of studies. As discussed earlier, these studies are divided into two main groups. The first focuses on what analysts say about the types of information they use in making their

[23] There are no written reports for buy-side financial analysts.
[24] Rogers and Grant (1997: 27) define comprehensive reports as reports with "complete firm reviews" while update reports "provide interim information about the firm or its activities".

decisions in response to questionnaires and/or interviews. The second analyses the content of analysts' reports.

The present study is based on content analysis of analysts' reports. This approach is motivated in part by Rogers and Grant (1997), who relate the information content of annual reports to the information content of analysts' reports. Rogers and Grant (1997) show that the information provided in the financial sentences represents a relatively small portion of the information provided in the analysts' reports. For the purpose of the current study, the most important finding in Rogers and Grant (1997) is: 'the MD&A section of the annual report is an extremely important section in terms of the information cited' (p.27). These findings support my decision to use analysts' reports in constructing the disclosure topics, and to use these topics to score annual report narratives.

4.2.4. The Content of Analysts' Reports

This section describes the structure of a typical comprehensive analyst report. Most reports start with an introduction. In this introduction, analysts discuss the principal rationales for their stock recommendations. Typically, the following points are discussed in the first two pages of any analyst report:

1) A short description of the company under investigation that includes its sector, market capitalisation, average daily volume, date of the report and the name of analysts and their address and contact number(s).
2) A target price for the common stock under evaluation, and a summary of the recommendation for the investment action (sell, buy, or hold).
3) The firm's product lines, its operations areas, its annual sales and profits, and its growth rate and its industry.

4) A summary of the company's historical trends in sales and earnings along with future prospectus.
5) A summary of the most recent significant developments in the firm's business (such as new products, mergers and acquisitions), industry competition or country of operation.

The report then produces a discussion to support the stock recommendation. Generally, analysts follow the top-down approach in writing their reports. This top-down approach is well explained by Hooke (1998). The approach consists of a number of subsequent levels of analysis. First, it investigates the principal economies in which a firm operates. It assesses the extent to which the state of economy can affect future share prices and industry earnings' growth. Second, it assesses the capital stock markets. Here, financial analysts are interested in examining the influence of market movements on share price. Third, analysts give a review of the company and its business, which includes industry analysis and firm specific analysis. In their industry analysis, analysts cover the major industry changes that are taking place and the new industry structures that are emerging. In their firm specific analysis, they cover topics related to a firm. These topics include products, customers, costs, sales and earnings. Fourth, any valuation models applied in valuing the firm are discussed. Finally, the report ends with a summary of the financial sentences, which includes the past, the current and the projected financial statements.

4.3. Overview of the Methodology

The purpose of the methodology adopted in this chapter is to discover the topics that help the market to forecast future earnings changes. Since the market is unobservable, it is assumed that analysts' reports can be used as a proxy for the stock market's view about the firm's disclosure quality. Such an assumption also underlies the use of AIMR-

FAF ratings as a proxy for the market's perception of the firm's disclosure quality. This is because analysts are widely considered to be a very influential user group in the market. They represent and influence investors' beliefs and activities (e.g., Schipper, 1991; Hirst et al., 1995 and Lang and Lundholm, 1996).[25] Hence, analysts' reports are text-searched for topics that are relevant in assessing a firm's future prospects.

The selection of disclosure topics is carried out in two main stages. The first stage involves creating a list of key words that are associated with forward-looking information in analysts' reports. In the second stage, the list of key words is used to identify sentences that are important to analysts in assessing a firm's future performance. For each forward-looking sentence, I identify the key noun of that sentence.

4.4. Analysts' Reports Dataset

This section describes the source of analysts' reports, the *Investext Plus* database. It also explains the criteria for selecting the sample of analysts' reports.

4.4.1. Data Source

The *Investext Plus* database is used to collect a sample of analysts' reports. Thomson Financial, the leading provider of e-information to the worldwide financial community, supplies this database.[26] It provides access to over 320,000 full-text company and industry research reports written by analysts at more than 270 of the leading investment banks, brokerage firms and consulting companies from 1996 onwards. These reports are displayed in PDF format - with images, charts, graphs, photographs and tables- appearing exactly as they were in a financial analyst's published documents.

[25] It is also argued that individual investors consider analysts' reports among the most influential sources for investment decision-making (SRI International, 1987).

[26] *Investext Plus* is used because this database contains a more comprehensive collection of analysts' reports than other databases such as *Dialog*.

4.4.2. Sample Selection Criteria

Due to the highly time consuming nature of this part of the study, as described later, it was decided to restrict the analysis to a relatively small sample of sell-side financial analysts' reports. The main sample is drawn from the UK brokerage houses. Year 1999 is selected to obtain a random sample of analysts' reports.[27] Analysts' reports are selected on the basis of:

1) They are written for the UK non-financial sector. This is because firms in the financial sector, such as banks, insurance and brokerage companies, have different activities and their reports are different from those of non-financial firms. As a result, the list of topics for financial firms may significantly differ from that of non-financial firms. Moreover, all types of non-financial firms are included in the study. This is because there is a need to construct a comprehensive list of topics published by analysts across industry sectors. This list will be used to measure the quality of disclosure for UK non-financial firms. Industry sectors are defined according to the classification given in the Financial Times in December 1999 (33 non-financial sectors);

2) They are written for the largest two firms ranked by market capitalisation for each sector. Exploring the availability of analysts' reports on the *Investext Plus* database for different firm sizes shows that financial analysts tend to write longer discussion sections for the largest firms.[28] Therefore, focusing on the largest firms in each sector increases the availability of analysts' reports with large discussion sections; and

3) The total number of pages for each report should be at least 20 pages. The sample is restricted to this number of pages because a comprehensive list of topics is needed to

[27] Year 1999 is the last year of the sample period used to undertake the regression analysis, but unlike previous years it does not suffer from a conversion problem that emerges when converting analysts' reports into text files.

[28] In the *Investext plus* database, I find that only firms in the top two deciles by size had analysts' reports with at least 20 pages. The only exceptions are some biotech firms. As a result, the selection of analyst reports with longer narrative portions is necessarily restricted to the largest firms by market capitalisation.

score annual reports. Longer analysts' reports might capture more topics than the smaller ones. If there is no an analyst report with 20 pages or more available for a firm, the next largest firm by market capitalisation is chosen as a replacement.

A list of analysts' reports with at least 20 pages for each firm is drawn up. Table 4.1 presents the *FT1999* industry classification and the selected firms in each sector. Table 4.1 illustrates that there are no analysts' reports in *Investext plus* with 20 pages or more for two sectors, which are 'Diversified Industries and Household & Textiles'. Only one analyst report with 20 pages or more is available for two sectors, which are 'Personal Care & Household Products and Steel & Other Metal'. The final list of analysts' reports represents 60 non-financial firms.

For many firms in the sample, there are several analysts' reports that are issued during 1999. However, one analyst's report is required for each of the firms in the current study. Therefore, a further three ad-hoc constraints are introduced to select the final sample of these reports. These constraints are discussed below.

First, the maximum number of reports for each brokerage firm should not be more than six reports. This represents 10 percent of the total sample. The rationale behind this constraint is that it prevents the list of topics from being dominated by one particular brokerage firm. In the present study, it is observed that some brokerage firms publish their reports with a standard style, with the same number of pages. They also tend to discuss the same topics for each company. For example, Merrill Lynch publishes a standard report with approximately 20 pages for any firm it covers.

In order to be sure that a wide range of topics are included in the scoring sheet, a second constraint is introduced. This constraint requires that two different brokerage firms should be selected for each industry sector.

Table 4.1. FT 12/1999 Industry Classification and the Selected Sample of the Largest Two Firms

Sectors	Selected Firms
1- AEROSPACE & DEFENCE	1- BRITISH AEROSPACE PLC. 2- ROLLS ROYCE PLC
2- AUTOMOBILE	1- GKN PLC 2- MAYFLOWER CORP.
3- BEVERAGE	1- DIAGEO PLC 2- ALLIED DOMECQ PLC
4- CHEMICALS	1- BOC GROUP 2- IMPERIAL CHEMICAL INDUSTRIES (ICI) PLC
5- CONSTRUCTION & BUILDING MATERIALS	1- HANSON PLC 2- BLUE CIRCLE INDUSTRIES
6- DISTRIBUTIONS	1- ELECTROCOMPONENTS 2- PREMIER FARNELL
7- DIVERSIFIED INDUSTRIES	NOT AVAILABLE
8- ELECTRICITY	1- SCOTTISH POWER PLC 2- NATIONAL GRID
9- ELECTRONIC & ELECTRICAL EQUIPMENT	1- BOWTHORPE PLC. 2- RACAL ELECTRICS PLC
10- ENGINEERING & MACHINERY	1- INVENSYS PLC 2- BBA GROUP PLC
11- FOOD & DRUG RETAILERS	1- TESCO PLC. 2- J. SAINSBURY PLC.
12- FOOD PRODUCERS & PROCESSORS	1- UNILEVER PLC. 2- CADBURY SCHWEPPES PLC
13- FORESTRY & PAPER	1- ARJO WIGGINS APPLETON 2- DAVID S. SMITH HOLDINGS
14- GAS DISTRIBUTION	1- BG PLC 2- CENTRICA PLC
15- GENERAL RETAILERS	1- KINGFISHER PLC 2- MARKS & SPENCER
16- HEALTH	1- NYCOMED AMERSHAM 2- SMITH & NEPHEW PLC
17- HOUSEHOLD & TEXTILES	NOT AVAILABLE
18- INFORMATION TECHNOLOGY HARDWARE	1- GENERAL ELECTRIC (GEC) PLC 2- ARM HOLDINGS
19- LEISURE, ENTERTAINMENT & HOTELS	1- GRANADA GROUP PLC 2- HILTON GROUP PLC
20- MEDIA & PHOTOGRAPHY	1- BRITISH SKY BROADCASTING 2- REUTERS GROUP
21- MINING	1- ANGLO AMERICAN PLC 2- RIO TINTO PLC
22- OIL & GAS	1- BP AMOCO PLC 2- SHELL TRANSPORT & TRADING
23- PACKAGING	1- REXAM PLC 2- MY HOLDINGS
24- PERSONAL CARE & HOUSEHOLD PRODUCTS	1- RECKITT & COLMAN PLC
25- PHARMACEUTICALS	1- GLAXO WELLCOME 2- ASTRAZENECA PLC
26- RESTAURANTS, PUBS & BREWERIES	1- BASS PLC 2- WHITBREAD PLC
27- SOFTWARE & COMPUTER SERVICES	1- SEMA GROUP PLC 2- MISYS
28- STEEL & OTHER METALS	1- CORUS GROUP PLC
29- SUPPORT SERVICES	1- RENTOKIL INITIAL PLC 2- CAPITA GROUP PLC
30- TELECOMMUNICATIONS SERVICES	1- VODAFONE AIRTOUCH PLC. 2- BRITISH TELECOM (BT).
31- TOBACCO	1- BRITISH AMERICAN TOBACCO PLC 2- IMPERIAL TOBACCO GROUP
32- TRANSPORT	1- RAILTRACK GROUP 2- BAA PLC
33- WATER	1- UNITED UTILITIES 2- THAMES WATER

Table 4.1 presents FT12/99 industry classification for UK non-financial firms. The largest two firms, according to market capitalisation, are selected. 'Household & Textiles' and 'Diversified Industries' have no analyst report with 20 pages or more. 'Steel & Other Metals' and 'Personal Care & Household Products' have only one firm with an analyst report of the required length.

Finally, it is intended to capture all types of analysts' reports. Therefore, the third constraint requires a selection of one comprehensive and one update report for each industry sector. This constraint ensures a selection of reports, which cover either the general review of the firm and/or those which cover special events. This constraint leads to a more comprehensive list of topics that captures different topics. For example, some annual reports disclosed information on special events such as mergers and acquisitions. As a result, it is logical to include these update reports as well as the comprehensive ones in the main sample.

To minimise the impact of the three ad-hoc constraints on the random sample, the selection process is started in industries with the smallest number of analysts' reports. Subject to these constraints, one analyst report is randomly selected for each of these 60 firms. The aggregate number of pages across the reports is 1898. This represents, on average, 32 pages in length per report, with the page length varying from a minimum of 20 pages to a total of 75 pages.

Table 4.2 provides the distribution of analysts' reports by brokerage houses. The total number of analysts' reports on the *Investext Plus* database across the 60 firms is 244. On average, four analysts' reports of 20 pages or more are available for the largest non-financial firms in the UK in 1999. Out of the 244 reports, 60 analysts' reports are randomly selected. These reports span 28 brokerage houses.

It is clear from Table 4.2 that the sample structure is a reasonable reflection of the population distribution. Only one brokerage firm 'Credit Lyonnais', is over-represented. This is due to the fact that the only reports available from 'Credit Lyonnais' are for two firms in the Water sector and, hence, these reports are automatically included in the sample.

Table 4.2. Topic Search: Distribution of Brokerage Houses

	Investext Plus		Sample	
Brokerage House	**Total**	**Percentage**	**Total**	**Percentage**
ABN AMRO BANK	25	10.2	6	10.0
ALBERT E. SHARP	2	0.8	2	3.3
ARNHOLD	1	0.4	0	0.0
BEAR, STEARNS	1	0.4	0	0.0
BNP EQUITIES	1	0.4	0	0.0
BT ALEX. BROWN	8	3.3	3	5.0
CCF ELYSEES	2	0.8	0	0.0
CHARTERHOUSE	5	2.0	1	1.7
CIBC WORLD	1	0.4	0	0.0
CREDIT LYONNAIS	10	4.1	6	10.0
CREDIT SUISSE	20	8.2	4	6.7
DEUTSCHE BANK	23	9.4	4	6.7
DONALDSON, LUFTKIN	8	3.3	2	3.3
DRESDNER KLEINWORT	23	9.4	5	8.3
HANDELSBANKEN	1	0.4	0	0.0
HSBC	7	2.9	2	3.3
ING BARING	1	0.4	0	0.0
LEHMAN BROTHERS	12	4.9	3	5.0
MERRILL LYNCH	22	9.0	6	10.0
MORGAN STANLEY	20	8.2	5	8.3
PARIBAS	8	3.3	3	5.0
SG COWEN SECURITIES	1	0.4	0	0.0
SG SECURITIES	13	5.3	2	3.3
THOMAS WEISEL	1	0.4	0	0.0
UBS WARBURG	5	2.0	0	0.0
WARBURG DILLION	17	7.0	6	10.0
WESTLB PANMURE	5	2.0	0	0.0
WILLIAMS DE BROE	1	0.4	0	0.0
Total	244	100.0	60	100.0

Table 4.2 shows the distribution of brokerage houses across 60 large UK firms in 1999. Columns 2 and 3 indicate the availability of analysts' reports with 20 pages or more on *Investext Plus*. For the topic search, one analyst report is selected per listed company. The resulting distribution is given in Columns 4 and 5.

Out of the 60 selected analysts' reports, 35 reports are classified as comprehensive, while the others are classified as updates. Comprehensive reports contain a complete discussion of a firm's activities. Update reports cover interim results (8 reports), general company updates (5 reports) and an analysis of special activities (12 reports).[29] Comprehensive reports dominate updates because update reports on the *Investext Plus* database in 1999 are not available for seven industry sectors.

[29] These activities are mergers, acquisitions, financing requirements, regulation issues, recommendation upgrades and business disposals.

Below, I describe the necessary procedure for constructing the scoring sheet. In addition, the extent to which *Nudist* assists in speeding up the process of creating such a sheet is also highlighted.

4.5. The Selection of Disclosure Items

Because of the matters raised earlier about the importance of forward-looking information, the focus of this chapter is to construct a list, which is sufficiently comprehensive to capture all forward-looking topics published in analysts' reports.[30]

The selection of disclosure topics is undertaken in two stages. The first stage generates a list of key words that are associated with forward-looking information in analysts' reports. In the second stage, this list is used to identify the topics that appear to be important to analysts in assessing a firm's future performance. For each forward-looking sentence, I identify the main topic of the discussion in that sentence. The following sub-sections discuss these stages in more details.

4.5.1. Selection of Forward-Looking Key words

Figure 4.1 summarises the steps followed in selecting the final list of key words. First, I read a sample of analysts' reports to identify the key words related to future events. This step produces a preliminary list of forward-looking key words. Second, synonyms for each key word are added to the preliminary list created in the previous step. Third, *Nudist* uses the new list of key words to text-search the 60 analysts' reports. Finally, two additional criteria are introduced that any key word must satisfy in order to be included in the final list. These are the frequency criterion and the forward-lookingness criterion.

[30] See Clarkson *et al.* (1999), Schleicher and Walker (1999), Miller and Piotroski (2000) and Hutton *et al.* (2002).

Below are the necessary steps to select the final list of forward-looking key words. These stages are: (1) reading analysts' reports, (2) adding synonyms & creating preliminary list, (3) identifying a list of forward-looking sentences, (4) testing the frequency of forward-looking key words, and (5) testing the forward-lookingness criterion.

(1) Reading analysts' reports

Creating forward-looking key words follows two stages. Firstly, a random sample of 30-analysts' reports from different brokerage firms published in 1998 is read and then a note is made for any key word that is related to future events. This process produces a preliminary list of forward-looking key words (list1).

(2) Adding synonyms & creating preliminary list

In this step I identify synonyms for the key words created in the previous step. Roget's Thesaurus is used to search for these synonyms. Then, these synonyms are added to the preliminary list (list1). A list of synonyms of the synonyms created in the last round is also added to list 1. The result is a new preliminary list (list 2) of 171 forward-looking key words (see Table 4.3).

Forward-looking key words in Table 4.3 can be classified into six groups. These groups are: 1) verbs such as 'anticipate', 'estimate', 'expect' and 'forecast', 2) nouns such as 'opportunity', 'prospect', 'outlook' and 'future', 3) adjectives such as 'new', 'next' and

Figure 4.1. Selecting the Final List of Forward-Looking Key words

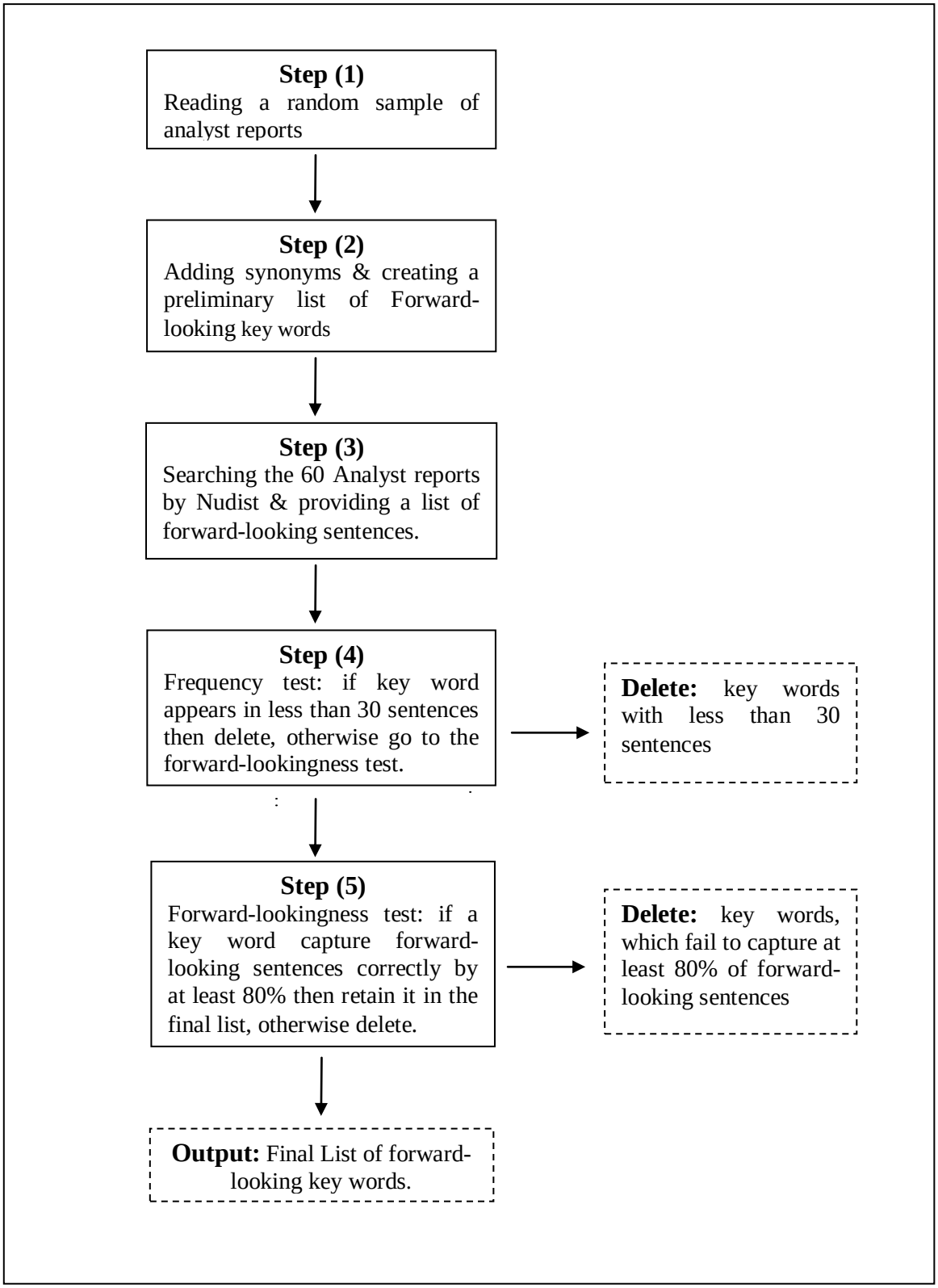


Table 4.3. The Preliminary List of Forward-Looking Key words

2000	Contemplate	Guess	Point toward	Strive
00E	Contingency	Hinder	Possibly	Subject to
00F	Continue	Hope	Potential	Suggest
Ability	Contract	Imagine	Precaution	Suppose
Certainty	Converted	Imminent	Predict	Surmise
Change	Convince	Impend	Presume	Target
Likely	Could	Improve	Presuppose	Turn
1999/2000	Decrease	Increase	Presurmise	Varies
Able	Deem	Innovation	Prevent	Vary
Accelerate	Designed for	Insight	Pro Forma	View
Additional	Designed to	Intend	Proceed	Vision
Advance	Desire	Intention	Program	Wait
Afterwards	Difference	Judge	Project	Well placed
Ahead	Divine	Keep	Prolong	Well positioned
Aim	Envisage	Later	Promise	
Allow	Envision	Likelihood	Prophesy	
Ambition	Estimate	Long for	Propose	
Anticipate	Eventual	Long-term	Prospect	
Approaching	Expand	Look	Purpose	
Approximate	Expansion	Maintain	Realise	
Aspire	Expect	Make larger	Reflect	
Assume	Extend	May	Remain	
Await	Extension	Medium-$term	Renew	
Become	Forecast	Might	Retain	
Beforehand	Forejudge	Model	Revitalise	
Belief	Foreknow	Near-term	Risk	
Believe	Foresee	New	Scenario	
Can	Foresight	Next	Scope	
Capable	Foretell	No later than	See coming	
Carry out	Forethought	Novel	Seek	
Challenge	Forthcoming	Objective	Shall	
Chance	Forward	Offer	Short-term	
Coming	Future	Opportunity	Shortly	
Commitment	Go faster	Optimistic	Should	
Committed	Go on	Outlook	Soon	
Confidence	Goal	Perspective	Speculate	
Conjecture	Going to	Pessimistic	Strategy	
Consider	Grow	Plan	Stretch	

Table 4.3 contains the preliminary list of key words that are extracted by reading sample analysts' reports and then using Thesaurus Dictionary for synonyms.

'novel', 4) adverbs like 'shortly' and 'soon', 5) date such as '2000', '2001'and '2002' and 6) phrase such as 'no later than', 'point toward' and 'willing to'.[31]

(3) Identifying a list of forward-looking sentences

In this step, the preliminary list of key words (list 2) is used to identify forward-looking sentences. First, I import analysts' reports into *Nudist*. Second, I use the 'text-search' function in *Nudist* to identify forward-looking sentences containing at least one of these key words. These two procedures are discussed in the following paragraphs.

[31] In some cases, I use $ at the end of the key words to instruct the programs to tag and count all word containing the term. For example, when one searches for the key word '$likely' via *Nudist*, the resulting node (file) contains all sentences with at least of the following key words 'likely', and 'unlikely'.

1- Importing analysts' reports into Nudist

Before introducing analysts' reports into *Nudist*, it is necessary to convert these reports into a standard text format and then to select a suitable text unit. These two requirements are discussed below.

a) Preparing analysts' reports

Documents are imported into *Nudist* as a standard text file. After selecting the sample of analysts' reports, the next task is to convert these reports from their original format (PDF format) to a standard text file format.[32] This conversion allows the reports to be imported into *Nudist*. Saving analysts' reports as text files removes special formatting such as bold, italics and centring. In addition, it deletes all images, charts, tables, pictures and graphics. But, information about the existence of these objects is kept in the text file.

The spelling of a random sample of sentences in these reports is also checked to be sure that there are no spelling mistakes after the conversion process. No non-traditional spellings or spelling mistakes are found in the selected sample. Finally, reviewing a sample of analysts' reports confirms that there is no missing information in these reports after the conversion process.

b) Selecting the text unit

Before importing analysts' reports into *Nudist*, a crucial stage is to decide the size of text units. A text unit is a portion of text that a researcher selects for the analysis. A text unit can be a paragraph, a sentence, a line, or a word. *Nudist* divides every document into a number of text units for the purpose of coding. In *Nudist*, one can use a

[32] Acrobat Reader version 5 is used in the present study to convert all analyst reports from PDF to standard text format.

paragraph, a sentence or a line as a text unit. The smallest text unit, 'word', is not included in the software.

The total number of text units in each document depends on the researcher's choice of text unit. This choice is a very significant decision and it will considerably affect the results of any research. Finally, it is not possible to change the selected text unit once the program is run.

A short paragraph may work well as a text unit, specifically if it contains one sentence. Practically, most documents have large paragraphs with many topics. Therefore, paragraphs are usually more difficult to code as a text unit than smaller portions, such as words, lines and sentences. Large text units typically contain more information and greater diversity of topics. Hence, they are more likely to provide researchers with conflicting signals. Here are two examples with *Nudist* when a paragraph is used as a text unit.

Example (1): A short paragraph

> *'The **acquisition** of Pioneer, announced on 29 November, makes Hanson the largest aggregates producer in the world and the second largest ready mix concrete producer'.* (Abn Amro, Hanson PLC, 1 December 1999, 2)

The paragraph above is acceptable as a text unit, because it is short and it includes only one clear topic 'acquisition'.

Example (2): A long paragraph

> *'As brokers to Hanson, we have withdrawn our **profit** forecasts and **investment recommendation**, pending completion of the Pioneer transaction. In the short term, Hanson has stated that the deal will be **earnings** enhancing in the first year pre-goodwill. **Cost savings** of at least £25m have been identified, but perhaps more interesting is the disparity in Pioneer's operating **returns** relative to Hanson, which could provide significant medium-term **benefits** to the group'.* (Abn Amro, Hanson PLC, 1 December 1999, 2)

The above paragraph is too long and it consists of many topics such as profit, investment recommendation, earnings, cost saving, returns and benefits. It also consists

of more than two sentences covering different tenses, which could give inaccurate results when using *Nudist* in identifying forward-looking disclosures.

A line, as a text unit, is better than a paragraph, but it does not work perfectly if the sentence is too long and consists of two or more lines. In this case, *Nudist* will not be able to locate phrases that extend across two lines or more. Below is an example to clarify the idea.

Example (3): Line as a text unit

> *'As brokers to Hanson, we have withdrawn our* ***profit*** *forecasts and'.* (Abn Amro, Hanson PLC, 1 December 1999, 2)

If a line is used as a text unit, it might be difficult to locate topics in one single line. The example above gives an incomplete picture about what an analyst is trying to argue in this sentence or paragraph. Only one topic, 'profit', appears but the overall meaning of the sentence or paragraph is incomplete.

Similar to Rogers and Grant (1997), a sentence is selected to be the basic coding unit in the present study. It is believed that a 'sentence' is more reliable than any other unit of analysis. The rationale for selecting this text unit is that coding text by sentence will provide results that better reflect the context of information that is discussed in the text.

In summary, a sentence is considered as the most appropriate information unit. Selecting sentences as a text unit changes the format of the analysts' reports. This is because each sentence will start with a new line after introducing these reports into Nudist. Sentences may be short or long and they may be clear or not clear. All these cases are discussed later in this chapter. Example (4) illustrates the use of a sentence as a text unit.

Example (4): A sentence as a text unit

> *'As brokers to Hanson, we have withdrawn our **profit** forecasts and **investment recommendation**, pending completion of the Pioneer **transaction**'.* (Abn Amro, Hanson PLC, 1 December 1999, 2)

The above example refers to one clear sentence, with one obvious tense. All topics also are clear and can be determined easily.

After preparing analysts' reports, the next procedure is to introduce these reports into a *Nudist* Project. There are two methods for importing documents into *Nudist*. In the first method, one can highlight *all his/her documents* then choose the *'Import Text Files as Documents'* function in *Nudist.* In the second method, one can use the command files.

In the present study, command files are used to import all analysts' reports. This is done by: (1) saving all analysts' reports in the Rawfiles folder in a *Nudist* project, and (2) running a command to import all these reports in a *Nudist* project. I use the following command to import the analysts' reports:

(introduce-documents all)

Using command files reduces the amount of time and effort to perform this type of routine task. Further benefits of using command files are discussed in the next chapter, when scoring large samples of annual reports.

2- Text-searching analysts' reports

In this stage, *Nudist* is used to text-search analysts' reports and produce lists of sentences that contain each forward-looking key word.

In the current procedure I benefit from different options in the 'text search' function. First, all analysts' reports are text-searched at one time. Second, the pattern search function is used which leads to searching for more than one key word at the same time

[e.g., (search-text $likely|forecast$|anticipat$)]. Finally, the whole-word search function is used in text-searching analysts' reports. For example, when analysts' reports are text-searched for the word 'will', the whole-word search function provides all sentences that contain the key word 'will'.[33]

The final outcome of text-searching analysts' reports is an initial sample of forward-looking sentences. These sentences are further analysed to select the final list of forward-looking key words.

Step (4) Frequency of forward-looking key words

This section introduces the first condition for selecting the final list of forward-looking key words. Each key word is used to text-search analysts' reports. For each key word, *Nudist* provides a statistic summary showing the frequency of each key word. This frequency is measured by adding up the total number of sentences associated with a single key word for each analyst report. Key words with less than a total number of 30 sentences are deleted.[34] All other sentences are kept for further analysis. The key words with 30 sentences or more are included in the final list of key words, only if they meet the forward-lookingness criterion.

Step (5) Forward-lookingness criterion

This criterion limits the number of key words to those that have a high probability of being associated with future events. In order to perform this test, a random sample of 30 sentences per key word is selected and read. Then, the number of cases in which the forward-looking key word is unambiguously associated with future events is counted. A

[33] If the whole-word function is not selected in this example, Nudist will provide all sentences that include the following key words, 'will, goodwill and/or willing'

[34] The rationales for deleting key words with less than 30 sentences can be summarised as follows. First, topics associated with most of these key words are either too general or too specific to industry sectors. Second, some of these key words usually introduce errors where they are in most cases associated with past sentences. Finally, a relatively large number of these key words are not common and contain topics that are discussed in the reports by using different forward-looking key words.

forward-looking key word is included in the final list if its associated sentences refer to the future in at least 24 cases (e.g., at least 80 percent of cases). The resulting list of 65 forward-looking key words is given in Table 4.4. Examples 6 and 7 clarify the idea.

Example (6): Forward-looking test 1

> **Sentence (1):** *'By year end we believe the **uncertainty will** have played out one way or the other'.* (Merrill Lynch, Anglo American PLC, 6 October 1999, 42)
> **Sentence (2):** *'**Uncertainty** on emerging markets **has raised** a high degree of forecast risk'.* (Merrill Lynch, Anglo American PLC, 6 October 1999, 58)

These two sentences illustrate how the forward-lookingness test works. In stage (4), the word 'uncertainty' is selected as a forward-looking key word. Text-searching analysts' reports provides a list of sentences associated with this word. On reading the first sentence, one would accept the word 'uncertainty' as a future key word. On the other hand, the same word in the second sentence is referring to the past. The inclusion of such a key word will introduce a noise when searching for the quantity of forward-looking information in annual reports. It will also introduce an error in scoring annual reports.

In the light of the above discussions, a forward-lookingness condition is informed to limit the number of key words that introduce noise. According to this condition, a forward-looking key word, such as uncertainty, should be included in the final list of key words if the sentences associated with it refer unambiguously to the future in at least 80% of cases. From the above two sentences, it is possible to argue that the word 'uncertainty' itself does not refer to the future. It does, however, refer to the future when accompanying another future key word like 'will' (see Sentence 1).

Table 4.4. Final list of Forward-Looking Key Words

Forward-Looking Key Word	Frequency	Forward-Looking Key Word	Frequency
2000	628	Outlook	159
00E	409	Plan	216
00F	71	Possible	171
Ability	105	Potential	415
Aim	87	Predict	40
Ambition	33	Prospect	201
Anticipate	103	Realise	109
Approximate	150	Reflect	330
Assume	657	Remain	359
Believe	693	Renew	35
Capable of	34	Risk	293
Chance	34	Scenario	70
Change	373	Scope for	72
Coming	84	Seek	64
Commitment	30	Short term	160
Continue	568	Soon	34
Could	598	Subject to	40
Estimate	1023	Suggest	284
Eventual	32	Target	477
Expect	1062	View	307
Forecast	640	Will	2088
Foresee	40	Would	1003
Forward	150		
Future	348		
Goal	31		
Going to	45		
Hope	45		
Improve	430		
Increase	645		
Intend	66		
Likely	585		
Long term	213		
Looking	114		
Looks	85		
Maintain	141		
Medium term	102		
Might	74		
Near term	45		
Next	478		
Objective	52		
Offer	561		
Opportunity	334		

Table 4.4 includes the 65 forward-looking key words that satisfied the frequency and the forward-lookingness tests. This list of key words is used to text-search the sample of 60 analysts' reports to identify forward-looking sentences.

Example 7 provides further explanations. In this example, the word 'may' or 'may be' are among the most popular forward-looking key words frequently used in analysts' reports. It provides a strong indicator of the future when it comes with another forward-looking key word as stated in the first sentence in Example 7. However, it sometimes

comes with past sentences (Sentence 2), or it can refer to the month 'May', not the verb 'may' (Sentence 3). In short, the key word 'may' does not satisfy the forward-lookingness condition because it introduces too much noise. Therefore, this key word is deleted from the final list.

Example (7): Forward-looking test 2

> **Sentence (1):** *'Oregon **is pursuing** deregulation and competition in the industrial sector, which **may be** introduced by 2001'.* (Donaldson, Scottish Power PLC, July 1999, 22)
>
> **Sentence (2):** *'This **has led** to the possible conclusion by the company that a sale* **may be** *the best way to realise value'.* (Credit Suisse, BAA PLC, 1999, 26)
>
> **Sentence (3):** *'The group entered the large Indonesian market, using Rothmans as a distributor, in **May 1999'**.* (Warburg Dillon, Imperial Tobacco PLC, 1999, 27)

4.5.2. Identifying Disclosure Topics

A topic is defined as the subject or the noun of a sentence. In this step, the nouns associated with the forward-looking key words are concerned. By using the 'Text Search' function in *Nudist*, all forward-looking sentences are identified in a very short period of time. Additionally, all other sentences either related to the past or any other irrelevant information are ignored.[35]

Forward-looking sentences are used to determine the topics to be included in the sheet.[36] To achieve this, the following stages are followed to construct the list of topics: (1) identifying the final list of forward-looking key words, (2) text-searching analysts' reports by the forward-looking key words, (3) creating a file contains all forward-looking sentences that contain at least one forward-looking key word, (4) reading *all sentences* for a set of forward-looking key words, (5) reading a random sample of

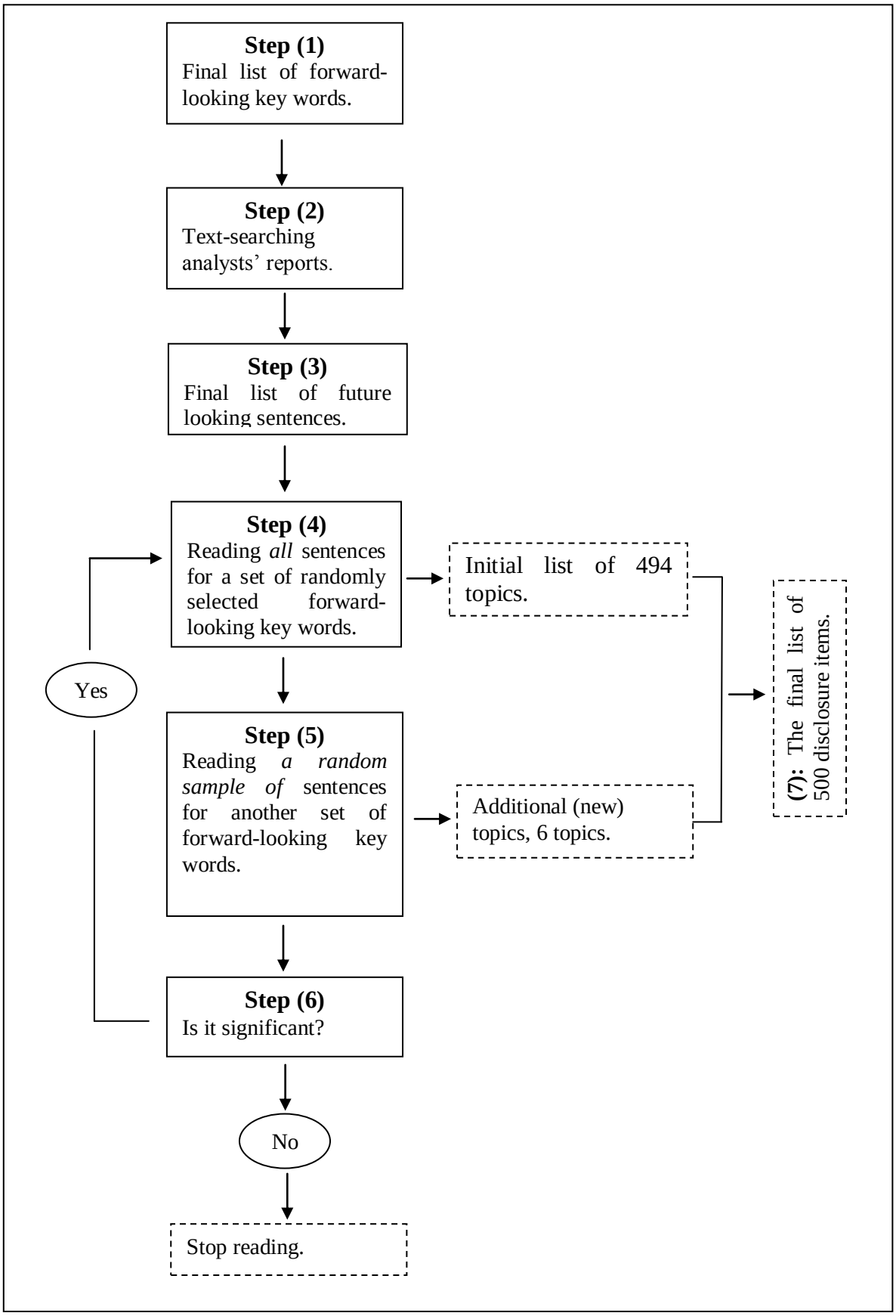
[35] Other irrelevant information refers to forward-looking sentences, which include an irrelevant forward-looking key word. Irrelevant key words are those which did not satisfy the forward-lookingness condition.

[36] An attempt has been made to automate the process of extracting topics from forward-looking sentences. However, automating such process was very difficult. Therefore, a sample of forward-looking sentences is read.

sentences for another set of forward-looking key words, (6) deciding the significance of the additional new topics and (7) creating the final list of topics.

Figure 4.2 provides a brief overview of these stages. Section 4.5.1 provides a detailed discussion regarding the selection of the final list of forward-looking key words. These key words are then used to identify future sentences in analysts' reports. The *Text Search* function in *Nudist* provides a great assistance in determining these sentences. The total number of forward-looking sentences for the 65 forward-looking key words is '20072'. On average, each forward-looking key word is associated with 309 sentences. It would have been too time-consuming to read all these sentences to identify the topics. Therefore, it was decided to identify the topics in a systematic way. Sixteen forward-looking key words are randomly selected. Then, *all sentences* containing at least one of these forward-looking key words are read. After that, another set of forward-looking key words is selected. For this new set, *a random sample of forward-looking sentences* containing at least one of these the key words is read. Finally, the percentage of the new additional topics is calculated. This percentage leads to make the final decision regarding the necessity of reading further forward-looking sentences. The following paragraphs provide a detailed discussion on the steps (4) to (7).

Figure 4.2. Selecting the Final List of Topics



Reading all sentence for a set of forward-looking key words

The final outcome of Section 4.5.1 is a final list of 65 forward-looking key words. This list is used to text-search a sample of 60 analysts' reports. This process produces a list of forward-looking sentences published in analysts' reports. In the current step, a random sample of sixteen forward-looking key words is read. Then, I decide to carefully read *all forward-looking sentences* containing at least one of these key words. The selected key words are shown in Table 4.5, Column 3. The total number of sentences read is 6639. This represents, on average, 415 sentences per forward-looking key word. This process produces an initial list of topics that are of interest to financial analysts in their forecasting process.

Forward-looking sentences that provide clear/specific information are focused on, while sentences that are more general or unclear are ignored. For each chosen sentence, a note is made for any forecasted noun(s) appearing in each sentence.

In many cases the forward-looking topic is obvious from a first reading. Example (8) illustrates these types of sentences.

Example 8: Identify disclosure topics

Sentence(1): *'We forecast **EBIT** to decline in FY1999 and again in FY2000 due to extremely weak prices'.* (Merrill Lynch, Anglo American PLC, 6 October 1999, 26)

Sentence(2): *'We believe that major **sales** are likely to arise in 2000 rather than this year'.* (Paribas, Glaxo Wellcome PLC, 16 June 1999, 16)

Sentence(3): *'A further £30m of **cost savings** are expected within the next three years'.* (Charterhouse Securities, J Sainsburys PLC, 14 July 1999, 19)

Sentence (4): '*The proportion of new **subscribers** appears to be increasing and churn rates are likely to fall resulting in strong growth in net new **subscribers** in 1999/00.*' (Dresdner Kleinwort Benson, British Sky Broadcasting PLC, 29 July 1999, 1)

Table 4.5. Reading Forward-Looking Sentences

Forward-Looking Key Word	Frequency	Reading Sentences	Forward-Looking Key Word	Frequency	Reading Sentences
Likely	585	All	Remain	359	Ignored
Expect	1062	All	Renew	35	Ignored
Forecast	640	All	Risk	293	Ignored
Estimate	1023	All	Scenario	70	Ignored
Anticipate	103	All	Scope for	72	Ignored
Will	2088	All	Seek	64	Ignored
Aim	87	All	Soon	34	Ignored
Hope	45	All	Subject to	40	Ignored
Intend	66	All	Suggest	284	Ignored
Long term	213	All	Target	477	Ignored
Medium term	102	All	View	307	Ignored
Near term	45	All	Would	1003	Ignored
Short term	160	All	2000	628	Ignored
Ability	105	All	00E	409	Ignored
Ambition	33	All	00F	71	Ignored
Approximate	150	All	Assume	657	Ignored
Capable of	34	Sample	Believe	693	Ignored
Chance	34	Sample	Could	598	Ignored
Change	373	Sample	Improve	430	Ignored
Coming	84	Sample	Increase	645	Ignored
Commitment	30	Sample	Offer	561	Ignored
Continue	568	Sample	Opportunity	334	Ignored
Eventual	32	Sample			
Foresee	40	Sample			
Forward	150	Sample			
Future	348	Sample			
Goal$	31	Ignored			
Going to	45	Ignored			
Looking	114	Ignored			
Looks	85	Ignored			
Maintain	141	Ignored			
Might	74	Ignored			
Next	478	Ignored			
Objective	52	Ignored			
Outlook	159	Ignored			
Plan	216	Ignored			
Possible	171	Ignored			
Potential	415	Ignored			
Predict	40	Ignored			
Prospect	201	Ignored			
Realise	109	Ignored			
Reflect	330	Ignored			

Table 4.5 includes the 65 forward-looking key words that satisfied the forward-lookingness test. *All sentences* associated with a random sample of these key words are read. A random sample of sentences associated with another set of these key words is also read. Finally, I decide to ignore the sentences associated with the remaining key words.

Example 8 shows that a financial analyst of Merrill Lynch (Sentence 1) provides a strong sentence about the direction of the current and the next year's EBIT, while a financial analyst of Paribas (Sentence 2) makes a fairly strong statement about the direction of sales in 2000. In Sentence 3, a financial analyst of Charterhouse makes a

strong statement about the expected cost savings within the next three years. In the final sentence, a financial analyst of Dresdner Kleinwort Benson provides a strong statement about the expected performance of British Sky Broadcasting in 1999/2000. The topics added to the list from the above sentences are 'EBIT', 'sales', 'cost savings' and 'subscribers'.

However, some sentences did not contain an obvious forward-looking topic, even though they contain one of the common forward-looking key words. Example (9) illustrates this case.

Example (9): Clear/Unclear topics

Sentence (1): *'Tesco believes that it should at least **aim** for something higher'.* (Morgan Stanley Dean Witter, Tesco PLC, 7 January 1999, 5).

Sentence (2): *'We believe that the company is **unlikely** to attain these expectations'.* (Paribas, Glaxo Wellcome PLC, 16 June 1999, 11).

Example (9) shows that the forward-looking key words 'aim' and 'unlikely' refer to the future. However, the sentences do not provide any specific topic.

The initial list of topics consists of 494 topics from reading forward-looking sentences that contain at least one of a random sample of forward-looking key words. It is important to note that some topics have more than one form, such as capital expenditure, cap. expenditure, cap. exp., and capex. Furthermore, some topics are presented as a single word and as a plural word such as (profit, profits). All these cases are aggregated into one topic.

Reading a random sample of sentences for another set of forward-looking key words

In this step, I randomly select another set of ten forward-looking key words. Then, I randomly select a sample of 50 sentences that are associated with these key words. These sentences are read to identify the percentage of new topics. The main objective of this stage is to know whether reading additional forward-looking sentences provides

additional topics to the initial list of topics created from the first sixteen key words. If so, is the total number of new additional topics statistically significant? In this stage, 420 sentences were read.[37]

Reading a further 420 sentences allows only six new topics to be added to the scoring sheet. Consequently, the most important question arising is: 'Will reading additional sentences significantly identify any new topics?' The following section provides an answer to the question.

Deciding the significance of the additional new topics

From the previous step, it seems reasonable to argue that reading additional sentences is not significant. Table 4.6 provides further evidence for this argument. Column 6 in Table 4.6 shows that the estimated loss of topics if *all* forward-looking sentences associated with the second set of selected key words is not read.[38] The estimated loss in this case is 33 topics. This represents only 1.95% (=33/1693) of the total number of sentences for these forward-looking key words. In short, it is possible to argue forward-looking key words that are selected in the first round capture most of the topics in analysts' reports.

[37] It is assumed that the total number of sentences read in this stage is 500 sentences (50 sentence* 10 forward-looking key words). However, some key words have less than 50 forward-looking sentences. Therefore, the total number of sentence equals to 420 sentences.
[38] Estimated loss of topics= % of new topics* frequency.

Table 4.6. Additional Disclosure Topics

Forward-looking key word	Frequency	Sample	New topics	% of New topics	Estimated loss
Capable of	34	34	1	0.029	1
Chance	34	34	1	0.029	1
Change	373	50	0	0.000	0
Coming	84	50	0	0.000	0
Commitment	30	30	1	0.033	1
Continue	568	50	2	0.040	23
Eventual	32	32	0	0.000	0
Foresee	40	40	0	0.000	0
Forward	150	50	0	0.000	0
Future	348	50	1	0.020	7
Total	1693	420	6	0.152	33

Table 4.6 provides the percentage of new topics added to the preliminary list if further forward-looking sentences are read. Estimated loss of topics is calculated as % of new topics* frequency

Creating the final list of topics

The new additional topics are added to the initial list. Therefore, the total number of items, which are finally selected, amounted to 500 topics. Table 4.7 provides a list of topics that comprise the scoring sheet.

From the scoring sheet, it can be seen that sell-side analysts produce forecasts in their reports in the following areas:

(1) Financial information: these types of information include:

1- Income statement information, such as income, profit, depreciation, revenue, costs and margins.

2- Balance sheet information, such as assets, liabilities, equity and capital.

3- Segmented information, such as lines, areas, businesses, divisions and subdivisions.

4- Cash flows information, such as cash inflows, cash outflows and cash flows from operations.

5- Financial ratios and valuation methods, such as CFPS (cash flows per share), CFROI (cash flows return on investment), COGS (cost of good sold), CROCI

(cash return on capital invested), EPS (earnings per share), DPS (dividend per share), EVA (economic value added), MVA (market value added) and FCF (free cash flows).

(2) Non-financial information: these types of information are divided into the following categories:

1- Company specific data, such as products, operations, management structure, strategy, product-life cycle and raw materials, production, distribution, suppliers, stores credit policy, customers, competition and merchandising.
2- Economic data, such as GDP, business cycle, inflation, consumption, tax rates and currency value.
3- Capital markets data, such as FTSE, interest rates, share price, ADR, market share, market size and market structure.
4- Industry analysis, such as technology, industry, demand, supply, competitors, government, rules, capacity and competition.

4.6. Evaluation

The new methodology introduced in the present chapter is clearly reproducible. This is because one could easily replicate all the process discussed in this chapter. The only needs are an electronic version of analysts' reports and a list of forward-looking key words. In the next chapter (Chapter 5), the necessary stages to score large samples of annual reports are discussed. This includes the selection of annual reports, the necessary procedure to automate the disclosure scoring process, the selection of the weighting methodology and the evaluation of *Nudist* as a tool for scoring annual reports.

Table 4.7. The Scoring Sheet

1	£/DM	51	Cash flow
2	£/Euro	52	Cash inflow
3	Accounting policy	53	Cash outflow
4	Accounting standard	54	Catalogue
5	Accounts payable	55	CFPS
6	Accounts receivable	56	CFROI
7	Accrual	57	Chain
8	Acquisition	58	Challenge
9	Adj. EPADR	59	Channel
10	Administration	60	Charge
11	ADR	61	Chemical
12	Advertiser	62	Chief executive
13	Advertising	63	Churn
14	Agreement	64	Client
15	Alliance	65	Closure
16	Ambitions	66	COGS
17	Amortisation	67	Combination
18	Approach	68	Commercial mileage
19	Approval	69	Commission
20	Arbitration	70	Committee
21	Area	71	Commodity
22	Asset	72	Compensation
23	Associated undertaking	73	Competition
24	Authority	74	Competitor
25	Backlog	75	Concession
26	Balance sheet	76	Conglomerate
27	Barrier	77	Consolidation
28	Benefit	78	Consortium
29	Beta	79	Construction
30	Bid	80	Consumer
31	Bill	81	Consumption
32	Bleak	82	Contract
33	Board	83	Contribution
34	Board meeting	84	Control
35	Bond	85	Convertible
36	Bond holder	86	Cost
37	Bonus	87	Coupon
38	Book value	88	Covenant
39	Borrowing	89	Cover
40	Brand	90	Coverage
41	Breakeven	91	Credibility
42	Budget	92	Credit
43	buy-back	93	Creditor
44	Buyer	94	Crisis
45	Capacity	95	CROCI
46	Capital expenditure	96	Crop
47	Capital	97	Culture
48	Capital structure	98	Currency
49	Capitalisation	99	Current account
50	Cash	100	Customer

101 Cycle
102 Cyclically
103 D/E
104 DACFPS
105 DCF
106 Deal
107 Debenture
108 Debt
109 Debtor
110 Deficit
111 De-listing
112 Delivery
113 Demand
114 De-merger
115 Deposit
116 Depreciation
117 Deregulation
118 Derivative
119 Design
120 De-stocking
121 Destruction
122 Devaluation
123 Developer
124 Development
125 Dilution
126 Director
127 Discount factor
128 Discount rate
129 Discovery
130 Disposal
131 Distribution
132 Distributor
133 Diversification
134 Divestment
135 Dividend
136 Division
137 DM/£
138 DM/Euro
139 Dollar
140 Downstream
141 Downturn
142 DPS
143 Drawback
144 Duty
145 Earnings
146 EBDIT
147 EBDITA
148 EBIT
149 EBITA
150 EBITDA
151 EBT
152 E-commerce
153 Economic turbulence
154 Economies of scale
155 Economy
156 Efficiency
157 Electricity
158 Employee
159 Engineering
160 Enhancement
161 Entity
162 Environment
163 EPADR
164 EPS
165 Equipment
166 Equity
167 EU
168 Euro
169 EV
170 EVA
171 Exceptional charge
172 Exceptional cost
173 Exceptional items
174 Exchange rate
175 Exit
176 Expansion
177 Expenditure
178 Expense
179 Exploration
180 Export
181 Exporter
182 Exposure
183 Extension
184 Facility
185 Factory
186 FCF
187 Federal reimbursement claim
188 Fee
189 Financing
190 Franchise
191 Franchisee
192 Franchising
193 FRS
194 FTSE
195 Fund
196 Gain
197 Gap between companies
198 Gas
199 GDP
200 Gearing
201 Geographic contiguity
202 Gold

203 Goods
204 Goodwill
205 Government
206 Grant
207 Growth
208 H1 results
209 Habit
210 Head office
211 Headcount
212 Hedging
213 Holding company
214 Import
215 Improvement
216 Incentive
217 Income
218 Indebtedness
219 Independence
220 Industry
221 Inflation
222 Infrastructure
223 Innovation
224 Instrument
225 Integration
226 Interest
227 Interim result
228 Internet
229 Inventory
230 Investment
231 Investor
232 IRR
233 IT
234 Job
235 Joint venture
236 Label
237 Labour
238 Land
239 Launch
240 Lawsuit
241 Layout
242 Leadership
243 Lease
244 Legal
245 Legislation
246 Lending
247 Leverage
248 Liability
249 Licence
250 Licensee
251 Liquidity
252 Listing
253 Litigation
254 Loan
255 Location
256 Loss
257 Machine
258 Machine line
259 Macroeconomic
260 Macro economy
261 Maintenance
262 Major force
263 Management
264 Manager
265 Manpower
266 Manufacturing
267 Margin
268 Market
269 Market position
270 Market price
271 Market share
272 Market size
273 Market structure
274 Market value
275 Marketing
276 Media
277 Merchandising
278 Merger
279 Mileage
280 Minority
281 Mix
282 MMC
283 Model
284 Momentum
285 Multimedia
286 NAV
287 Net present value (NPV)
288 Network
289 NOPAT
290 NOPLAT
291 NPAT
292 NWC
293 Obligation
294 Occupancy rates
295 Offering
296 Oil
297 OPEC
298 OpFCF
299 Operation
300 Operational benefits
301 Operator
302 Option
303 Order
304 Order-book

305 Ordinary shares
306 Organisational structure
307 Outlet
308 Out performance
309 Output
310 Outsourcing
311 Overcapacity
312 Overdraft
313 Overhead
314 Packaging
315 Partners
316 Partnership
317 Patent
318 pay-back
319 Payment
320 Payout
321 Pay-to-basic ratio
322 PBIT
323 PBT
324 PE
325 Penetration
326 Performance
327 Perpetuity
328 Personnel
329 Phase
330 Plant
331 Platinum
332 Players
333 Political
334 Politician
335 Population
336 Position
337 Preference share
338 Premium
339 Presence
340 Price control
341 Price cut
342 Price inflation
343 Price limits
344 Pricing
345 Privatisation
346 Proceed
347 Processor
348 Producer
349 Product
350 Product line
351 Production
352 Productivity
353 Profit
354 Profitability
355 Programming
356 Progress
357 Project
358 Promotion
359 Property
360 Provision
361 PTP
362 Purchasing
363 PV
364 R&D
365 Raw material
366 Reappraisal
367 Recession
368 Recovery
369 Redundancy
370 Refurbishment
371 Region
372 Regulation
373 Regulator
374 Regulatory
375 Reinvestment
376 Remuneration
377 Renewal
378 Rent
379 Reorganisation
380 Replacement
381 Repurchase
382 Reputation
383 Research
384 Reserve
385 Residual value
386 Resource
387 Restructuring
388 Result
389 Retailer
390 Retailing
391 Retiring
392 Return
393 Revaluation
394 Revenue
395 Reward
396 Rights
397 Risk
398 ROA
399 ROACE
400 ROCE
401 ROE
402 ROI
403 ROIC
404 Royalty
405 Ruling
406 Rural

407	Salary	454	Tariff
408	Sale	455	Tax
409	Sales	456	Technical
410	Saving	457	Technological
411	SEC	458	Technology
412	Sector	459	Territory
413	Security	460	The world's top company
414	Segment	461	Threat
415	Seller	462	Tooling
416	Selling	463	Top line
417	Senior Executive team	464	Trading
418	Service	465	Traffic
419	Setback	466	Training
420	Settlement	467	Translation
421	Share	468	Trend
422	Share issue	469	Trial
423	Share price	470	Turbulence
424	Shareholder	471	Turmoil
425	Shareholding	472	Turn
426	Shopper	473	Turnaround
427	Site	474	Turnover
428	Slippage	475	UK GAAP
429	Slow down	476	Uncertainty
430	Slump	477	Underachievement
431	Social security	478	Underperformance
432	Software	479	Unit
433	Space	480	Upgrade
434	Speciality	481	Upgrading
435	Spending	482	Uplift
436	Spin off	483	Upstream
437	Square feet	484	Upturn
438	Staff	485	User
439	Stage	486	Utilisation
440	Stake	487	Valuation
441	Standard	488	Value added
442	Start up	489	VAT
443	Sterling	490	Venture
444	Stock	491	Visitor
445	Stockholding	492	Volume
446	Store	493	WACC
447	Store openings	494	Wage
448	Strategy	495	Waste
449	Subscriber	496	White paper
450	Supplier	497	Workforce
451	Supply	498	Working capital
452	Synergy	499	Worth
453	Takeover	500	Yield

Table 4.7 provides a list of disclosure items (nouns and phrases) that are published in a random sample of analysts' reports. Reading a sample of forward-looking sentences creates these items. Unclear, firm-specific and double entries items are deleted. The remaining items are ordered. For each disclosure item, single and plural topics are aggregated into topic. Also, a disclosure item with more than one format is aggregated into one topic.

4.7. Summary

A criticism of the disclosure index studies stated by Schadewitz and Blevines (1997) in their review article is that these studies focus on annual reports and ignore any other media of communication. This chapter provides a new methodology for constructing the list of disclosure items. Novel features of this methodology are related to the use of analysts' reports to construct the list of disclosure items and the ability to use computer software to speed up the process of constructing this list.

A random sample of 60 analysts' reports is used to identify the forward-looking topics that are used by financial analysts. 35 of these reports are comprehensive while the others are updates. The reports of 28 different UK brokerage firms are included in the sample.

To select the disclosure items, a sample of analysts' reports is read to identify a preliminary list of forward-looking key words. Then, the synonyms for all these key words are identified. After that, two additional criteria are introduced, which any key word must satisfy in order to be included in the final list. These are the frequency criterion and the forward-lookingness criterion. The final list of forward-looking key words is used to collect the forward-looking sentences appearing in 60 analysts' reports. The text-search function in *Nudist* achieves this task. Finally, a sample of forward-looking sentences is read to identify the topics in each sentence. The final scoring sheet comprises a list of 500 topics that appear to be important for financial analysts in their forecasting process.

Chapter 5: A New Methodology for Scoring Annual Reports

5.1. Overview

This chapter presents a new methodology for scoring annual reports. An important feature of this methodology is that it automates the generation of disclosure scores through the use of *Nudist*. This automation allows me to produce disclosure scores for large samples of firms at minimal cost. The resulting sample sizes are comparable to those employed by US researchers with access to annual AIMR-FAF disclosure ratings.

The work on analysts' reports, described in Chapter 4, produces the set of forward-looking topics that analysts frequently refer in their reports (List A). I use List A to score annual report narratives. Before I can use this topic list, however, I also need a set of forward-looking key words similar to the set of forward-looking words that I use in the analysis of the analysts' reports. Since it is possible that firms may use different forms of expression to refer to the future, I decide to construct a new set of forward-looking key words from a sample of 50 annual reports (List B). Hence I produce two lists for the automated scoring of the annual report. List A contains a set of topics that analysts frequently refer to in their reports. List B contains forward-looking words such as 'expect', 'anticipate' and 'predict'. I score annual reports via *Nudist* to automatically count the number of sentences in the relevant sections of the annual report that contain at least one forward-looking word and at least one relevant topic.

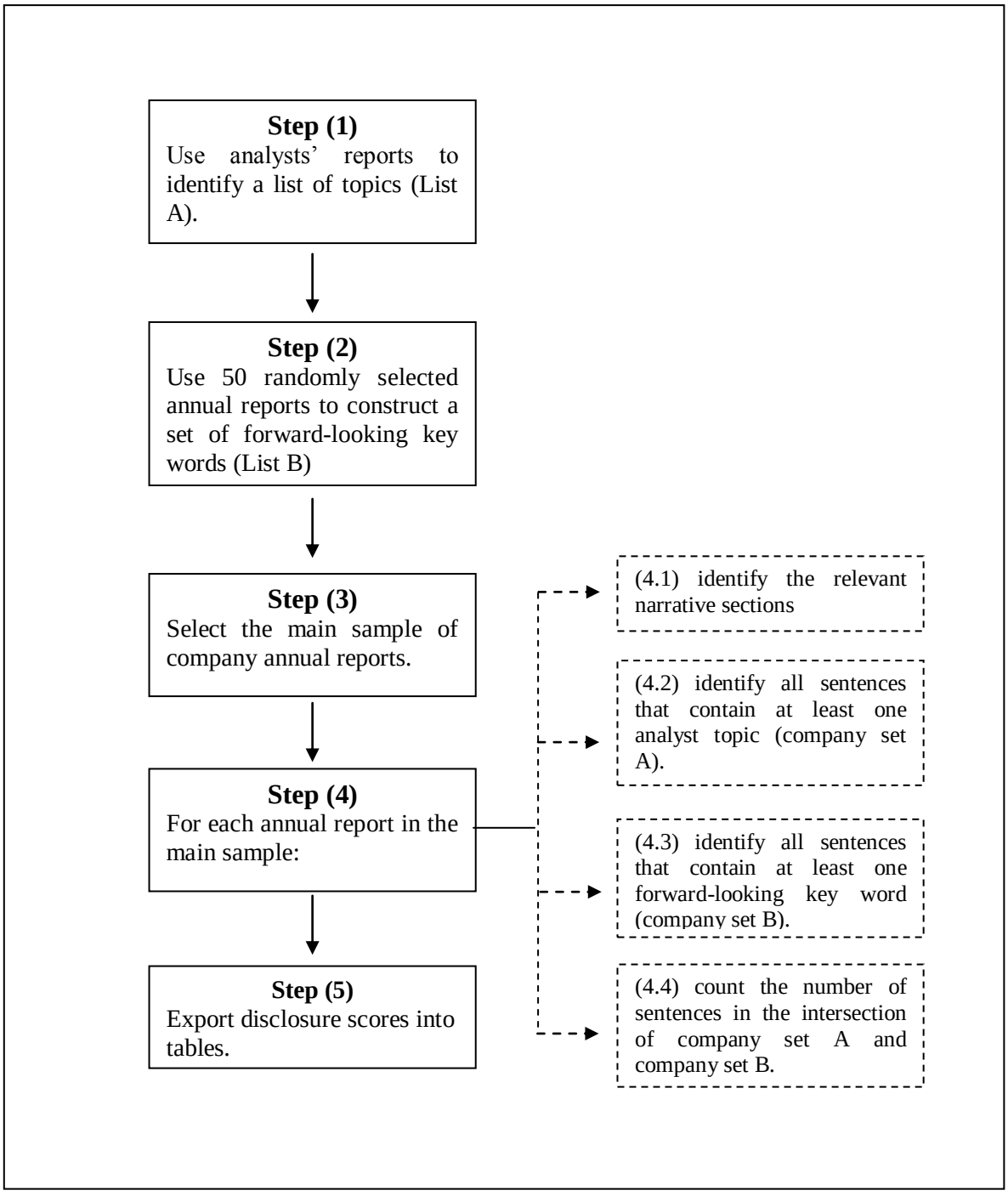
The reminder of this chapter is organised as follows. Section 5.2 presents the scoring procedures. Section 5.3 discusses issues related to the reliability and validity of the methodology. An evaluation of the scoring methodology is described in Section 5.4. Section 5.5 comments on the scoring methodology adopted in the present study. Section 5.6 summarises.

5.2. Scoring Annual Reports

The scoring of annual report narratives is the most crucial part of this study. Figure 5.1 describes the steps that are followed to score the sample of annual reports. Scoring annual report narratives is carried out in five steps. First, analysts' reports are used to identify a list of topics (List A). This is the list of disclosure topics contained in Table 4.7. Second, a random sample of 50 annual reports is read to construct a set of forward-looking key words commonly employed by companies (List B). The final list of forward-looking key words is shown in Table 5.1. Third, the main sample of company annual reports is selected. Fourth, for each annual report in the main sample I (1) identify the relevant narrative sections, (2) find the set of sentences containing at least one key word from List A to produce company annual report set A, (3) find the set of sentences containing at least one key word from List B to produce company annual report set B, and (4) automatically count the number of sentences in the intersection of set A and set B. Finally, *Nudist* is used to export the number of sentences in the intersection of set A and set B into tables.

In scoring annual report narratives, the following conventions are employed: (1) all sentences in the intersection of company set A and company set B are counted regardless of the narrative section of the annual report in which they are found, (2) all selected items are considered relevant and material, and (3) the borderline between past and future is the financial year-end.

Figure 5.1. Scoring Annual Reports



The simplest kind of nominal scale, a Dichotomous Scale, is chosen to evaluate annual report narratives. This scale provides only two categories, which are disclosed or not disclosed. An item is regarded as disclosed (not disclosed), then the company scores one (zero). The disclosure score for each annual report is additive and unweighted. Using an unweighted scoring technique assumes that each item is of equal importance. By using this technique, the subjectivity involved in assigning weights to the different items, when the user's preferences are unknown, is reduced (see Gray et al., 1995).

The following sections describe the steps that are executed for automating the generation of disclosure scores for large samples of annual report narratives.

5.2.1. The Construction of Disclosure Topics

The work on analysts' reports, described in Chapter 4, produces the set of forward-looking topics that I use to score annual report narratives. The total number of topics generated by this stage is 500. This topic list is referred as List A.

5.2.2. The Construction of Forward-Looking Key Words

The process of identifying forward-looking key words used in the annual report is similar to the process adopted when I identify forward-looking key words used in the analysts' reports (see Section 4.5.1). In short, I follow three stages to select the final list of forward-looking key words. These stages are discussed in the following paragraphs.

In the first stage, a random sample of 50 annual reports is selected and read. A note is then made for any forward-looking key word that appears. For each key word, Roget's Thesaurus Dictionary is used to search for synonyms. For each synonym, any further synonyms are identified and then are added to the list created in the previous phase. This stage ends by producing a preliminary list of forward-looking key words similar to those shown in Table 4.3.

In the second stage, a sample of 788 annual reports in 1999 is text-searched.[39,40] Searching these reports produces a list of sentences that are related to the key words created in the first stage.

The final stage is to identify the final list of forward-looking key words. To produce this list, a random sample of 30 sentences per forward-looking key word is selected. These sentences are read. The total number of sentences that clearly referred to the future is then calculated. Forward-looking key words are included in the final sample if their associated sentences refer to the future in at least 67% of the sentences that are read.[41] In addition to the forward-lookingness test, further analysis is undertaken. The aim of this analysis is to examine the extent to which forward-looking key words are used accurately in annual reports. In analysts' reports, the main focus is on disclosure topics that analysts try to forecast in their reports. It does not matter for creating disclosure topics if they refer to the past, the current, or the future prediction. However, for annual reports, the situation is different. This is because the main objective of creating the scoring sheet is to link the quality of forward-looking disclosure to share price anticipation of earnings. As a result, selecting the correct forward-looking key words is crucial for scoring annual reports. The following example clarifies that idea. The following sentences are produced by a text search of four annual reports for the key word 'anticipate':

[39] 800 annual reports for 1999 are randomly selected from *Dialog*. 12 firms are deleted from this sample because of the unavailability of *Datastream* codes. Further discussion is presented in the next section.
[40] Before importing these reports into Nudist, a 'sentence' is selected to be the text unit (see more discussion in Chapter 4).
[41] Here I introduce a subjective benchmark in determining the final list of forward-looking key words. The reasons for choosing this cut-off are: (1) the need of choosing less noisy forward-looking key words to score annual reports. I think that the higher the benchmark, the lower the number of noisy forward-looking key words to be included in the final list and (2) the need to provide researchers with clear instructions of the procedures I used to replicate my work in the future and to communicate the reliability of my disclosures scores.

Example: Further analysis

> **Sentence (1):** '*Continental Europe exhibited some slowdown during the second half and we **anticipate** that this is likely to continue to be the pattern for the majority of the coming year'.* (Acal plc, 1999 annual report and accounts)
>
> **Sentence (2):** '*The Cardowan development site has progressed well and it **is anticipated** that the infrastructure works will be completed by early summer'.* (Alexander Russell plc, 1999 annual report and accounts)
>
> **Sentence (3):** '*Souvenir/Ritepoint, our writing instrument business, improved sales and earnings marginally on 1997 but not to the levels we **had anticipated'**.* (4imprint plc, 1999 annual report and accounts)
>
> *Sentence (4): '**As anticipated**, sales to the UK Public Sector and Government continued to decline'.* (AEA technology plc, 1999 annual report and accounts)

Searching for the key word like 'anticipate' produces a list of sentences that refer to the future as in Sentences 1 and 2. This frequently occurs, specifically when the sentences include further key words such as 'likely' (in Sentence 1) and 'will' (in Sentence 2). On the other hand, it also introduces some noise as shown in Sentences 3 and 4. Both of these sentences include backward looking information. This noisiness occurs in several verbs such as forecast, expect, estimate and predict. Therefore, an attempt is made to reduce such noise as discussed in the next paragraph.

To reduce the noise induced by verbal key words, key words with more specific tense structure are used to text-search annual reports. For example, instead of using 'anticipate' as a key word, different forms of this key word are used. These forms are 'anticipate', 'anticipates', 'is anticipated', 'are anticipated', 'is anticipating' and 'are anticipating'. Narrowing the search in this way helps to avoid the past tenses related key words that appear when one text-searches annual reports for the key word 'anticipate' (e.g., 'have anticipated', 'was anticipated' and 'had anticipated'). Therefore, the noise in verbal key words is reduced.

The final list of key words is given in Table 5.1. This table shows the importance of each key word. Column 2 represents the total number of the selected forward-looking

Table 5.1. Forward-Looking Sentences: List of Key Words

Key Word	Forward-Looking Sentences	Hits	Reports
Accelerate	23	472	226
Anticipate	20	1045	475
Await	25	43	37
Coming [financial] year[s]	27	326	204
Coming months	25	34	31
Confidence, Confident	23	1238	543
Convince	25	44	40
Current [financial] year	29	905	407
Envisage	27	80	68
Estimate	20	596	266
Eventual	22	55	49
Expect	29	1826	537
Forecast	20	366	219
Forthcoming	24	120	101
Hope	24	169	137
Intend, Intention	27	763	387
Likely, Unlikely	27	396	232
Look ahead, Look forward	25	400	271
Next	24	1100	455
Novel	23	131	41
Optimistic	26	62	58
Outlook	24	505	350
Planned, Planning	30	206	141
Predict	30	61	51
Prospect	25	1073	515
Remain	20	1826	601
Renew	20	287	156
Scope for, Scope to	28	96	75
Shall	30	163	123
Shortly	25	173	130
Should	22	1097	480
Soon	28	119	101
Will	30	11226	772
Well placed, Well positioned	30	192	149
Year[s] ahead	30	153	107
1999/2000, 1999–2000	27	170	89
2000, 2001, 2002 … 2009	28	1471	324

Table 5.1 contains the list of key words that are used to identify forward-looking sentences in the annual report discussion section. Column 2 measures the success of a key word in separating forward-looking from backward-looking information. The number of forward-looking sentences relates to a sample of 30 randomly selected sentences. Columns 3 and 4 indicate the total number of sentences ('Hits') and the total number of discussion sections ('Reports') that contain the key word. Both statistics relate to 788 annual reports from the year 1999. If applicable, descriptive statistics are calculated after excluding from the search certain endings that are frequently associated with backward-looking sentences. '1999/2000' and '1999–2000' refer to a firm's next financial year. The year numbers '2000' to '2009' must be preceded by one of the following prepositions: 'during', 'for', 'in', 'into', 'of', 'through' or 'throughout'.

sentences for each key word.[42] Column 3 shows the frequency of each key word. Column 4 represents the total number of 1999 annual reports containing the key words.

[42] This number is related to a randomly selected sample of 30 sentences.

It is clear from Column 3 that the key word, 'will', dominates all other key words in terms of both hits and frequency.[43]

5.2.3. The Main Sample of Annual Reports

5.2.3.1. Data Source

As mentioned before, *Nudist* is used to score large samples of annual reports. To be analysed, these reports must be machine-readable (stored in an electronic form). Recently, the availability of UK annual reports in an electronic form has been improved.[44] Thus, UK researchers have the opportunity to select large samples of annual reports and, hence, to undertake large-scale disclosure studies.

This study uses *Dialog* to collect annual reports. Dialog is a relatively new database. It is the most comprehensive source of UK annual reports.[45] This database maintains a collection of the annual reports from approximately 2500 UK firms from 1995 onwards[46]. These reports are provided in a standard text format.[47] Therefore, the conversion problems that are associated with the other sources of annual reports are avoided .[48]

---

[43] Hits refer to the total number of future sentences for each key word, while the frequency refers to the total number of 1999 annual reports containing the key words.

[44] In the UK, different databases now contain large samples of annual reports in an electronic format. These databases include *Dialog*, *Perfect Information*, *Carol* and *Northcote*.

[45] At an early stage of the study, I check the accuracy of annual report information supplied by *Dialog*. In particular, I compare sales figures on *Datastream* for 100 firms with those figures published in annual reports. No errors are found. I also compare a sample of electronic annual reports with their hard copies, and again I found no mistakes or errors.

[46] *Dialog* also maintains a collection of half yearly interim reports and a collection of analysts' reports.

[47] *Dialog* deletes any images, graphs, pictures, tables and charts in the annual report, but it retains all text and numbers. Therefore, the number of pages for each report in *Dialog* should be smaller than that of its original hard copy.

[48] Perfect Information, Carol and Northcote databases provide annual reports with PDF format. At an early stage, a problem was encountered in converting annual reports to text format when Perfect information was used.

5.2.3.2. Sample Selection Criteria

For a firm to be included in the sample, it should satisfy the following conditions:

1- It should belong to a non-financial sector. Financial firms such as banks, insurance firms and investment firms are excluded. This is because their reports are not comparable to those of non-financial firms.

2- For each non-financial firm, one annual report, at least, should be available on *Dialog* from 1996-1999. This study focuses on annual reports, not other media of financial communications such as interim reports.

The first year on *Dialog* is 1995. Unfortunately, the number of annual reports for 1995 is rather small in comparison with the other years, so it is not included in my sample. Therefore, the sample of this study starts in 1996. 1999 is chosen as the end year for the study. This is because, in the empirical part of the study, the quality of corporate disclosure is linked with share price anticipation of earnings. Consequently, accounting and return data is required for at least two years ahead (Year 2001). Annual reports after 1999 are ignored due to the unavailability of accounting and return data for two years ahead at the time of undertaking the analysis. Thus, the final sample period covers 1996 to1999.

The total number of firms on *Dialo*g for the years 1996 to 1999 is 1594, 2005, 1947 and 1892. After deleting financial companies, this reduces to 1122, 1410, 1368 and 1289. *Dialog* covers 1740 non-financial firms during 1996-1999. The total number of annual reports for these firms in these periods is 5189. Some firms change their financial year-end and, hence, they produce two annual reports in one specific year. In *Dialog*, there

are 19 non-financial firms with two annual reports. As a result, the total number of annual reports for non-financial firms in 1996-1999 is 5208.[49]

The above paragraph shows that the total number of firms varies from year to year. For example, the total number of firms in 1996 is 1594. This number increases in 1997 to 2005 firms, and then it is reduced to 1947 firms in 1998 and to 1892 firms in 1999. It is not clear why *Dialog* covers some firms while it ignores others.

Due to the time consuming nature of collecting annual reports from Dialog, saving them and deleting irrelevant sections as discussed later, I decided to randomly select only 800 annual reports for each year. This represents a total of 3213 annual reports (note that some firms have two annual reports). In detail, the total number of annual reports for each year is: 803 annual reports for 1996, 803 annual reports for 1997, 801 annual reports for 1998 and 804 annual reports for 1999.

The selected companies are matched with the *Datastream* codes. Some firms have no *Datastream* code. Thus, they have no accounting and return data. These firms are excluded from the selected sample. Accordingly, the sample is reduced from 3213 to 3142 annual reports. This represents 783 reports in 1996, 781 reports in 1997, 790 reports in 1998 and 788 reports in 1999. The resulting panel includes 1558 companies. The average number of annual reports per company is 2.02.

[49] The total number of annual reports for non-financial firms = [1122+1410+1368+1289] +19= 5208.

5.2.4. Producing Disclosure Scores

5.2.4.1. Identifying the Relevant Narrative Sections

This study focuses on voluntary disclosures in the annual report discussion section. This is because these sections are more likely to include forward-looking information than the financial sentences.

Following Beattie et al. (2002), the text-search analysis is based on sections that contain the following (or similar) headings in the annual report: Financial Highlights, Summary Results, Chairman's Statement, Chief Executive Officer's Review, Operating and Financial Review, Financial Review, Financial Director's Report, Finance Review, Business Review, and Operating Review.

The other sections of the annual report are deleted before importing these reports into *Nudist*. These sections include the following (or similar) headings: Table of Contents, List of Directors, Report on Corporate Governance, Statement on Environmental Policies, Statement on Health and Safety Policies, Directors' Report, Report from the Remuneration Committee, Statement of Directors' Responsibilities, Auditors' Report, Financial Sentences, Notes to the Financial Sentences, List of Principal Operating Companies, Historical Summary, and Shareholder Information. Such sections are either largely standardised or contain information that mainly serve a stewardship function.[50]

5.2.4.2. Identifying all Sentences Containing at Least One Analyst Topic

Using *Nudist*, all annual report narratives are searched for sentences that contain at least one topic in List A. The text-Search function in *Nudist* is a useful tool for this task. It provides considerable assistance in identifying the required sentences for large samples

[50] The process of collecting annual reports from Dialog, saving them on a computer and deleting the irrelevant sections, takes around ten minutes per report. As a result, about eight weeks was needed to collect annual report narratives from *Dialog*, change the firm's name with their *Datastream* codes and to prepare the reports for the analysis.

of companies. Command files are used to speed up the process of identifying the forward-looking sentences.

The study uses different functions to text-search annual reports by using *Nudist*. First, Pattern-Search function is used to search for a group of key words in one command file. For example, it is used to text-search annual reports for the key words ‘profit’, ‘profitability’, ‘profitable’ and unprofitable’. In addition, Whole Word Only function is usually selected to force *Nudist* to only select sentences containing the exact key word. For example, when text-searching annual reports for the key word ‘Profit’ and restricting the search to the whole word only, the output will be the sentences associated with the exact key word ‘profit’. However, if the whole word function is not chosen, the results will be all sentences which include the following key words ‘profit’, ‘profits’, ‘profitable’, and ‘unprofitable’.

The next example illustrates the way in which I use the command files to perform a search for the key word, ‘likely’:

Example: The Use of Command Files

(search-text ‘cost’ pattern-search? no whole-word? no case-sensitive? no node (9 1) node-title ‘cost$’)	
Command lines	**Comment**
(	Start command file
Search text	Start a text search
‘cost’	Search for the key word ‘cost’
Pattern search? No	Do not use the pattern search
Whole-word? No	Do not search for the whole word only
Case sensitive? No	Do not require case sensitivity
Node (9 1)	Save found sentences in node 9 1
Node title ‘cost$’	Name node 9 1 ‘cost$’
(	End command file

In the above command file, pattern search is not used because only one key word ‘cost’ is searched for. In addition, the whole word option is not used in the search because

both the key words, ‘cost’ and ‘costs’, are searched for.[51] The result of text-searching process for the key word ‘cost’ is given in free node (9 1).

A number of command files are written similar to the above example, to text-search annual reports for the 500 disclosure topics included in Table 4.7. These commands are then run for each individual year of the study.[52] The final outcome from this analysis is a list of sentences containing at least one disclosure topic for each firm in the sample year.

5.2.4.3. Identifying all Sentences Containing at least One Forward-Looking Key Word

The work of this section is similar to the one described in the previous section. However, instead of using analyst topic key words, this stage uses the set of forward-looking key words that are included in Table 5.1 to text-search annual report narratives. The final outcome from this analysis is a set of sentences that contain at least one forward-looking key word from list B for each firm in the sample year.

5.2.4.4. Identifying all Sentences Containing at least One Forward-Looking Key Word and One Analyst Topic

Nudist is used to identify the set of sentences that contain at least one analyst topic and at least one forward-looking key word (the intersection of company annual report set A and company annual report set B).

Section 5.2.4.2 shows the set of sentences in an annual report that contains a relevant topic. Section 5.2.4.3 shows the number of sentences in an annual report that deals with forward-looking sentences. An advanced feature of *Nudist* allows me to find the

[51] The ‘$’ sign does not allow Nudist to search for the whole word.
[52] Technically, *Nudist* treats each year as a specific project.

intersection of these two sets. The disclosure score is the number of sentences in this intersection.

5.2.5. Exporting the Results into Tables

A particularly useful feature of *Nudist* is that it allows me to export these disclosure scores directly into a spreadsheet. This spreadsheet shows the *Datastream* code and the total number of forward-looking sentences with a relevant topic for each firm.[53] These scores can then be combined with the firm's accounting and return data using standard statistical software such as SAS.

5.3. Assessment of Reliability and Validity of Disclosure Scores

For a valid inference, Weber (1990) argues that the classification procedure should be reliable and valid. Reliability refers to the ability of different people to code the same text in the same way (Weber 1990:12). Validity refers to the extent to which the variables generated from the classification procedure represent what the researcher intends it to represent.

5.3.1. Assessment of Reliability of Disclosure Scores

In the present study, all annual reports are coded at one time. After a short period of time, samples of these reports are coded again. The resulting scores yielded from the second time phase coincide exactly with those arrived at the first time round. This provides assurance of stability of the coding methodology.

Marston and Shrives (1991) argue that the index scores awarded to companies could be considered reliable if other researchers could replicate (reproduce) the same results. To improve the reliability of disclosure scores in the present study, a clear statement of

[53] A *Datastream* code is used (rather than the firm's name) to identify each annual report analysed by *Nudist*.

procedures is developed and a clear definition of forward-looking information is adopted. Therefore, other researchers should be able to replicate the disclosure scores reported on this study. Figure 5.1 provides the stages that are followed to score a large sample of annual reports. In summary, in order to replicate the new scoring methodology introduced in this study, one needs two lists: (1) a forward-looking key words list, and (2) a disclosure topics list.

5.3.2. Assessment of Validity of Disclosure Scores

Validity refers to the extent to which the variables generated from the classification procedure represent what the researcher intends it to represent. The validity of these disclosure scores is supported by three different sets of analyses: (1) the correlation between the number of forward-looking profit sentences for each firm and the total number of forward-looking sentences for the same firm, (2) the correlation between disclosure scores and firm characteristics identified in prior studies to be associated with the level of corporate disclosures, and (3) the correlation between the automated disclosure scores and the scores calculated by reading annual report narratives.

An examination of the internal consistency of the disclosure index provides some insights into the validity of the disclosure scores. In particular, since corporate disclosure activities are coordinated across various avenues, I predict that each of the components of the disclosure index will be positively correlated with each of the remaining components and the total disclosure scores. In the present study, the analyses are based on two sets of disclosure scores. The first is defined in terms of an all-inclusive topics list, while the second defines disclosure quality in terms of forward-looking profit topics only. The Pearson correlation between these disclosure scores is positively significant, equals 0.76, at the 0.0001 level.

The second type of analysis is based on the correlation between disclosure scores and firm characteristics identified in prior studies to be associated with the level of corporate disclosures. Ahmed and Courtis (1999) offer a meta-analysis of the results of 23 separate studies of the association between the levels of disclosure in the annual report and firm characteristics since 1961. The findings of their study show that only four variables have a significant positive association with disclosure levels. These are: firm size, exchange listing status, audit firm size and leverage.

In the present study, I focus only on firm size for a number of reasons: (1) audit firm size is not *electronically* available at the time of undertaking the analyses, (2) the sample is based on all UK listed companies. In addition, a number of studies still cannot find a significant association between disclosure levels and listing status (e.g., Botosan, 1997), and (3) there is evidence that the predicted association between the levels of disclosure and leverage does not exist (Hail, 2002).

Because of the above reasons, I decide to focus on the size variable. Prior studies find the predicted association between the levels of disclosure and firm size. In the present study, the Pearson correlation between disclosure scores and firm size is positively significant, equals 0.48, at the 0.001 level for an all-inclusive topics list, and it is also positive and significant, equals 0.26, at the 0.001 level for a profit topics list.

The final type of analysis calculates the correlation between the automated disclosure methodology and the traditional methodology that is based on reading annual reports. In order to undertake this analysis, a sample of annual reports is read in order to calculate the frequency of the forward-looking sentences. For the same sample of reports, *Nudist* is used to calculate the frequency of forward-looking sentences. Then, the results yielded by *Nudist* are compared against those yielded by reading annual reports. The

test yields the ratio of forward-looking statements correctly identified by *Nudist* to all forward-looking sentences. This ratio is calculated as [1587/(1287+1587)]=55.2 percent. The corresponding value for backward-looking sentences is given by [7376/(7376+233)]=96.9 percent. This result supports the validity of my automated disclosure scores. The following sections provide more details regarding this final test.

Taken together, the validity of the disclosure scores is generally supported by the above three sets of analysis.

5.4. Evaluation

This section examines the extent to which the use of key words in *Nudist* effectively differentiates between forward-looking information and backward-looking information. This is done by comparing and contrasting the differences between two methods used in scoring annual reports. The first is based on *Nudist* (computerised content analysis) and the second is based on reading annual reports (traditional content analysis). Therefore, the research questions under this section of study are:

1- To what extent do the traditional and the computerised approaches yield similar results with regard to forward-looking information and backward-looking information?
2- In what ways do the results differ and what are the reasons behind these differences?
3- What is the correlation between the results of both methods?
4- What are the benefits and limitations of both methods?

In order to be sure that the new scoring methodology adopted in the present study produces reasonable results, 50 annual report discussion sections from the year 1999 are randomly selected. Each discussion section is read and I identify sentences about the

future. I then automate the search by using *Nudist* and the key words in Table 5.1 and compare the two classifications. The result of this cross-tabulation is given in Table 5.2.[54]

Table 5.2. Evaluating Disclosure Scores: Forward-Looking Versus Backward-Looking Sentences

	Researcher Forward-Looking	**Researcher Backward-Looking**
Nudist – Backward-Looking	1287 (12.3)	7376 (70.4)
Nudist – Forward-Looking	1587 (15.1)	233 (2.2)

Table 5.2 provides evidence on the success of the key words in separating forward-looking from backward-looking sentences. I carefully read the discussion section of 50 annual reports from the Year 1999 and separate information into forward-looking and backward-looking sentences. I then automate the search by using *Nudist* and the key words in Table 5.1 and compare the two classifications. The total number of sentences in the 50 annual reports is 10483. Numbers in parentheses indicate percentages.

Table 5.2 shows that the number of sentences identified by *Nudist* correctly is calculated by [70.4+15.1] is 85.5 percent. This is a very encouraging result, which reflects the success of the key words combined with *Nudist* to score annual reports accurately. This result suggests that *Nudist*'s text-searches tools are a reliable alternative to other scoring methodologies. However, one should interpret this percent carefully. This is because another way to analyse Table 5.2 is to calculate the ratio of forward-looking statements correctly identified by Nudist to all forward-looking sentences. This ratio calculated as [1587/(1287+1587)] is 55.2 percent. The corresponding value for backward-looking sentences calculated as [7376/(7376+233)] is 96.9 percent.

Table 5.2 also shows that 14.5 percent are misclassified, which represents 12.3 % for error type I (forward-looking information is identified by manual reading, but not by Nudist) and 2.2 % for error type II (forward-looking information is identified by *Nudist*, but not by manual reading). Most of these errors happen when *Nudist* classifies

[54] In very few cases it is ambiguous as to whether the sentence is forward-looking or backward-looking. In these cases one of my supervisors is consulted to obtain a second independent opinion.

forward-looking sentences as backward-looking sentences. One-way to reduce error type I is to lower the benchmark for including key words in Table 5.1. However, this reduction would come at the cost of increasing the type II error. The next paragraph provides a further discussion of these types of errors.

The ratio of forward-looking statements correctly identified by Nudist to all forward-looking sentences is relatively low (55.2%). This means that a number of forward-looking disclosures *are omitted* or *are not captured* by *Nudist*. The reason for that can be summarised as follows:

1) Some key words are not included in the final sample of forward-looking key words because of the benchmark for including key words in Table 5.1 (67%). Examples of these key words are: may be, would be, can be, continue, strategy, propose, recommend, aim, purpose and objective. These types of key words introduce a considerable noise and subjectivity when determining the number of forward-looking sentences in the annual report electronically.

2) The difficulty of identifying the future sentences that include the following format: 'be'+ 'verb' + 'ing'. For example, the Chairman of the company can refer to the improvement of the quality of customer service by the following sentence: we **are making** effort to provide a good service to our customers on time. Unfortunately, *Nudist* cannot categorise this sentence as a forward-looking sentence.

3) Some backward-looking sentences carry messages which have relevance for the future. For example, the announcement of backward-looking information such as an increase in capital expenditure or the R&D during the last year has value-relevant information for investors and analysts in predicting future corporate performance (see, for example, Bryan, 1997).

In addition, the low ratio of forward-looking statements correctly identified by Nudist to all forward-looking sentences suggests that *Nudist* categorises some backward-looking information as forward-looking information. In other words, there are some sentences captured by *Nudist* as forward-looking disclosures while they are omitted when I use the manual reading. The reasons for this case can be summarised as follows:

1) Some forward-looking key words such as (2000) introduce relatively minor errors. For example, when one searches for the 1999 annual reports and uses the year '2000' as a key word, *Nudist* cannot recognise the difference between 2000 as a forward-looking key word and 2000 as a part of date written after the chairman's signature.

2) Sometimes in annual reports, one paragraph starts with a title such as 'Outlook or Forecasts'. In the next line, the chairman or the CEO starts the first sentence of his/her paragraph with one sentence as an introduction to past events. Then in the second sentence, he/she introduces his/her forecasts for the following year(s). *Nudist* considers the title of the paragraph as a part of the first sentence. As a result, this sentence counts as a forward-looking sentence, though it may refer to the past. The following example illustrates this case.

> ***PROSPECTS***
> *'In 1997 our ports business began to show encouraging results from the capital investment programme undertaken in recent years'.* (Associated British Ports Holdings PLC- 1997 Annual Report and Accounts)

In the above example, the sentence clearly refers to a past event, but *Nudist* scores it as a forward-looking sentence. This problem cannot be avoided when a sentence as a text unit is used. Actually, this problem represents a slight drawback of the *Nudist* program. However, the very significant gains in efficiency are established when using *Nudist* compared with the risk of these errors.

In order to test *Nudist*'s ability to identify forward-looking sentences and to score firms correctly, two statistics are required. These statistics are: (1) the correlation between the number of forward-looking sentences identified by reading the reports and the number of those sentences identified by *Nudist*, (2) the percentage of sentences detected by reading the reports that is correctly identified by *Nudist*.

For the first test, the Pearson's and Spearman's Rank correlations are used to assess the strength and the direction of the relationship between the two methodologies. A strong positive Pearson correlation (0.96, significant at the 0.001 level) between the numbers of forward-looking sentences identified by reading the reports and those identified by *Nudist* is found. Also, a strong positive rank correlation (0.95, significant at the 0.001 level) is found between the two methodologies. These results indicate that the scores resulting from *Nudist* are closely correlated with those resulting from reading annual reports.

The second test evaluates *Nudist*'s ability to correctly identify forward-looking sentences. The total number of forward-looking sentences identified by *Nudist* is 1820. Reading annual reports identifies 1587 of these sentences correctly. This means that 85.5% of forward-looking sentences identified by *Nudist* are correct.

In summary, this section has sought to make a comparison between the two methods of scoring annual reports. These methods are the traditional scoring method and the computerised one. In the traditional method, all sentences are read. For each sentence, one can identify whether a sentence includes forward-looking or backward-looking information. In the computerised method, a list of forward-looking key words is used to differentiate between forward-looking and backward-looking sentences. The overall results show that there is a high correlation between the two methods. Therefore, it is

possible to argue that the *Nudist* program works successfully in the sense that very important gains in efficiency are recognized compared with the risk of errors.

5.5. Comments on the Scoring Methodology

The new scoring methodology presented in this chapter has several advantages relative to the labour-intensive scoring methods. These advantages include: (1) the ability to automatically score very large numbers of annual reports at very low marginal costs, (2) an increase in the comparability of the disclosure scores across firms and over time, and (3) the ability to replicate the scores easily in subsequent disclosure studies.

In addition, the new scoring methodology offers advantages relative to the subjective ratings such as AIMR-FAF ratings. These advantages include:

1) The ability of the researcher to identify his/her own criteria for the particular requirements of his/her research project. In the present study, the emphasis is on forward-looking information.

2) The ability of the researcher to control the sample selection criteria and sample period. This is because the main data requirement for firms to be included in the dataset is the availability of annual reports in an electronic form.

3) Other researchers can replicate disclosure scores easily. To replicate the scoring system, one only needs two lists of key words and topics.

On the other hand, the new scoring methodology has potential limitations. These include:

1) It calculates disclosure scores by adding up the number of forward-looking sentences for each report. This technique ignores the fact that the usefulness of disclosures can vary from sentence to sentence,

2) It also ignores firm characteristics like number of segments, industry membership, etc. when calculating disclosure scores. Such characteristics are potentially important inputs when assessing the adequacy of disclosures, and

3) It equates disclosure quality with the amount of information provided. It ignores the underlying tone of forward-looking sentences. Identifying the underlying tone of voluntary disclosures means that one can discriminate between good news and bad news. Such a refinement could be useful when examining the economic consequences of corporate voluntary disclosures.

If the limitations of the new scoring system are important empirically, then they will reduce the power of the tests and will work against my ability to find the predicted association between disclosure quality and 'prices leading earnings'.

5.6. Summary

This chapter is in the spirit of Core's (2001) call for research designed to assess the quality of corporate disclosure. As he discussed 'Improved measures of disclosure quality also need to be developed. The AIMR discontinued its disclosure rankings in 1997 (after ranking fiscal year 1995). There may be some small problems of judgement error in the metrics constructed by Botosan (1997), Lang and Lundholm (2001) and Miller (1999), but the real problem with these measures is that they are so labour-intensive that they are feasible only for small samples. Here, I conjecture that researchers can substantially lower the cost of computing these metrics by importing techniques in natural language processing […]'. In this spirit, the main purpose of this chapter was to introduce a new methodology for scoring a large number of annual reports.

Five stages are followed to score corporate annual reports. First, analysts' reports are used to identify a list of disclosure topics. Second, I select a random sample of 50 annual reports to construct a set of forward-looking key words. Third, the main sample of company annual reports is selected. Fourth, for each annual report in the main sample I (1) identify the relevant narrative sections, (2) find the set of sentences containing at least one forward-looking key word, (3) find the set of sentences containing at least one disclosure key word, and (4) automatically count the number of sentences that are forward-looking in nature and contain at least one relevant topic. Finally, the overall results of the scoring methodology are exported to a spreadsheet.

The reliability and the validity of my disclosure scores are supported by different sets of analyses. These include (1) a clear statement of procedures that are followed to score annual reports, (2) the correlation between the number of forward-looking profit sentences for each firm and the total number of forward-looking sentences for the same firm, (3) the correlation between disclosure scores and firm's size, and (4) the correlation between the computerised disclosure scores and the scores calculated by reading annual report narratives.

Nudist is used to perform the descriptive work in this study. Using this software helps the researcher in creating the disclosure index. It also provides assistance in automating the generation of disclosure scores for a large sample of annual reports. The total number of future sentences counted by *Nudist* is highly correlated with the number resulting from manual scoring. Therefore, it would appear that *Nudist* works successfully in the logic that the extremely significant gains in efficiency are established in relation to the risk of minor errors.

Chapter 6: Disclosure Quality and Prices Leading Earnings: A Cross-Sectional Analysis

6.1. Overview

A major contribution of Chapters 4 and 5 is the development of a new scoring methodology that partially automates the generation of disclosure scores for large samples of firms. In the present chapter, these disclosure scores are used to examine one particular disclosure issue. Following Lundholm and Myers (2002) and Gelb and Zarowin (2002), I investigate whether higher levels of disclosure increase the strength of the relation between current stock returns and future earnings changes. Unlike Lundholm and Myers (2002) and Gelb and Zarowin (2002), however, I emphasise individual classes of information and also the sensitivity of the findings to the precise way in which forward-looking information is identified.

Following Collins et al. (1994), I regress current returns on current and future earnings variables. Then, I allow the regression coefficients to vary with the firm's disclosure quality. The empirical results provide evidence that forward-looking profit information helps the market to anticipate future earnings changes more accurately. The results of this chapter are important because they suggest that the scoring methodology developed in Chapters 4 and 5 appears to work well in a cross-sectional analysis.

The remainder of this chapter is organised as follows. Section 6.2 presents the regression model and the hypotheses. Section 6.3 describes the data. Section 6.4 provides the definition for the variables used in the study. Section 6.5 presents the descriptive statistics and the Pearson correlations of the variables used in the study. The main empirical results are presented in Section 6.6. Section 6.7 summarises.

6.2. Research Design

As discussed earlier, the framework adopted in this study is the one introduced by Collins et al. (1994). This model is given as:

$$R_t = b_0 + b_1 X_t + \sum_{k=1}^{N} b_{k+1} X_{t+k} + \sum_{k=1}^{N} b_{k+N+1} R_{t+k} + b_{2N+2} EP_{t-1} + b_{2N+3} AG_t \qquad (1)$$

Where:

R_t : Stock return for period t.

X_t : The growth rate of earnings for period t.

EP_{t-1} : The earnings yield for period t–1.

AG_t : Asset growth for period t.

R_{t+k} : Stock return for period t+k.

X_{t+k} = The growth rate of earnings for period t+k .

As discussed in Chapter 3, Collins et al. (1994) include future earnings growth variables into the return-earnings regression model. Using the Collins et al. (1994) model, Lundholm and Myers (2002) argue that it is possible to detect the source of changes in expected future earnings. More specifically, they argue that corporate disclosure is a significant source of information about a company's future performance.

In order to examine the effect of disclosure on prices leading earnings, I interact all right-hand side variables in equation (1) with a dummy variable, D , where D is defined to be one for high disclosure firms and zero for low disclosure firms.[55] This is consistent

[55] A dummy variable approach has the advantage of providing a test statistic that indicates whether the difference in the coefficient estimates between two groups of observations is statistically significant. The number of dummy variables is an empirical choice. Choosing a single dummy variable approach allows me to use a maximum number of observations for the estimation of the interaction term coefficients.

with the approach in Lundholm and Myers (2002) and Gelb and Zarowin (2002). Interacting all explanatory variables with D yields the main regression model in this chapter:

$$R_t = b_0 + b_1 X_t + \sum_{k=1}^{2} b_{k+1} X_{t+k} + \sum_{k=1}^{2} b_{k+3} R_{t+k} + b_6 EP_{t-1} + b_7 AG_t + \mathrm{b}_8 D + b_9 [D * X_t] + \sum_{k=1}^{2} b_{k+9} [D * X_{t+k}] + \sum_{k=1}^{2} b_{k+11} [D * R_{t+k}] + b_{14} [D * EP_{t-1}] + b_{15} [D * AG_t] + e_t \quad (2)$$

Two changes are made relative to the Collins et al. (1994) regression model. First, only two future earnings growth variables are incorporated in the regression model (N = 2 and k = 1, 2) rather than three future years as in Collins et al. (1994). This is because, at the time of the data collection, most of the firms from the year 1999 had no earnings data for year t+3. Second, earnings growth variables are defined by deflating earnings change by stock price and not by lagged earnings. This is due to the fact that it is difficult to define earnings growth when lagged earnings are negative. Therefore, the use of an earnings deflator would reduce the sample sizes further. As a result, a price deflator is used instead of the earnings deflator.

The theory underlying model (2) implies a number of predictions for the signs of the model parameters. More specifically, b_1 is expected to be positive (see Lev, 1989). Also, the future ERCs of low disclosure firms, b_2 and b_3, are expected to be positive. This is due to the fact that industry-wide and economy-wide effects should allow the market to anticipate some portion of the firm's future earnings change, even if the annual report discussion sections do not include forward-looking information.

b_4 and b_5 are expected to be negative. This is because any unanticipated future events lead to higher earnings growth in the period t+k should also lead to positive returns in

the period when the news becomes available to the market. Hence, there is a positive relation between unanticipated future earnings and future returns. The coefficient b_6 is expected to be positive. This is due to the fact that a high stock price in relation to last year's earnings signals high expected earnings growth for the current and future years. This implies a negative association between earnings yield and expected earnings growth and, hence, the coefficient on EP_{t-1} is positive. b_7 is predicted to be negative because of the positive association between asset growth and the expected growth in future earnings. There is no particular prediction for b_8.

Since the level of corporate disclosure may cause a substitution effect away from current earnings and toward future earnings, Lundholm and Myers (2002) predict a negative coefficient on $D * X_t$. On the other hand, additional voluntary disclosure might make earnings announcements more credible. Hence, the sign for b_9 is difficult to predict. The coefficients b_{10} and b_{11} are predicted to be positive. This is because the relation between stock returns and future earnings changes should increase with disclosure quality, and b_{10} and b_{11} measure the difference in future ERCs between high and low disclosure firms.[56] Finally, there are no particular predictions for the coefficients b_{12}, b_{13}, b_{14} and b_{15}.

Following Lundholm and Myers (2002) and Gelb and Zarowin (2002), the present study seeks to test the hypothesis that forward-looking information published in annual report narratives leads to a significant improvement in investors' forecasts of future earnings growth. More specifically, if a firm reveals forward-looking information in their annual report, then this information should be reflected in market prices. As a result, one would

[56] Note that the future ERC of high disclosure firms is obtained as the sum of b_{k+1} and b_{k+9}.

expect that realised future earnings will be partially anticipated by current stock returns. If this is the case, then the coefficient on interacted future earnings, $D * X_{t+1}$ and $D * X_{t+2}$, will be positive in the return-earnings regression model (2). In other words, high levels of forward-looking disclosure reveal information about future earnings.

In summary, the above arguments indicate that there is an interactive effect between the levels of forward-looking disclosure in annual report narratives and future earnings growth. Thus, it seems reasonable to hypothesise that:

H1: High levels of forward-looking information in annual report narratives are associated with stock prices that are more informative about future earnings growth.

Hypothesis 1 is tested for an all-inclusive topic list as identified in Section 4.5. This list of topics is given in Table 4.6. While an all-inclusive topic list might be a fair reflection of the forward-looking discussion in analysts' reports, it is clear that many included topics are at best weakly correlated with future earnings changes. For example, while forecasts of future debt levels have implications for interest charges and hence bottom-line profits, other factors like interest rate cuts and rises obviously disturb the direct relation between debt and interest expense. Such topics might introduce noise when examining the effect of disclosure on prices leading earnings. Accordingly, I also decide to use a narrow definition of disclosure quality. This definition is based on a new list of topics that are more likely to have a strong effect on the return-earnings association.

The new list of topics is a sub-sample of topics identified from analysts' reports, but it contains only topics that I judge to be related directly to profits. This list of topics is given in Table 6.1.

Table 6.1. List of Profit Topics

Topic	Synonyms and Related Topics
Profit	Benefit, Breakeven, Budget, Contribution, Earnings, EPS, Loss, Margin, Profitability, Return, Trading

Table 6.1 contains 12 topics that are closely related to profit. The list of topics in Table 6.1 is a subsample of topics identified from analysts' reports. Where applicable, the plural of a topic is also included in the text search.

Choosing a profit topic list is motivated by the fact that this list contains information that is more likely to be correlated with future earnings changes. In addition, US researchers provide evidence that these forecasts are both informative and credible. They find that management earnings forecasts influence stock prices, alter investors' earnings expectations, improve analyst forecast accuracy, reduce analyst dispersion and lower bid-ask spreads (see discussion in Piotroski, 2002). Disclosure scores based on profit topics are used in the analysis to test the second hypothesis, which is:

H2: High levels of forward-looking profit information in annual report narratives are associated with stock prices that are more informative about future earnings growth.

6.3. Data

The sample period comprises the years 1996 to 1999. The total number of non-financial firms on *Dialog* for these years is 1,122, 1,410, 1,368 and 1,289, respectively. From these I randomly select 800 annual reports per year. This provides a sample of 3,200 firm-year observations.

50 firms are deleted from the sample due to the unavailability of *Datastream* codes. Consequently, the annual reports with *Datastream* codes are 3,150 firm-years. The resulting sample panel contains 1,558 firms. The average number of annual reports per firm is 2.02.

Before conducting the descriptive analysis, calculating the correlation matrix and executing the regression analysis, further observations are deleted. First, any firm that changes its accounting year-end during the period 1996–1999 is deleted. Second, missing observations are deleted. Third, outliers are censored to avoid any undue influence of extreme observations. Outliers are defined in the present study as the top and bottom 1% of observations for the distribution of any of the regression variables.[57] Finally, a large number of observations are deleted because of the definition given to high and low disclosure firms. In particular, disclosure scores for the whole sample are divided into four quartiles. The second and third quartiles are deleted. This represents 641 firms in panel A and 671 firms in panel B. Firms in the first quartile are defined as high disclosure firms, while firms in the fourth quartile are defined as low disclosure firms.[58] This leaves 917 (for the all-inclusive topic list) and 887 (profit topics only) firm-year observations from the period 1996 to 1999. Note that the number of usable observations varies with the definition of disclosure quality.

6.4. Variable Definition

The association between disclosure quality and the return-earnings association is tested using Collins et al. (1994). In this model, current stock returns is the dependent variable, while the independent variables include current and future earnings, future returns, earnings yield and asset growth. Earnings and accounting data are collected from *Datastream*.[59] The definition of each variable is discussed in the next paragraphs.

---

[57] Excluding extreme observations is consistent with prior literature (e.g., Kothari and Zimmerman, 1995). In addition, censoring the top and bottom 1% of observations is one of the acceptable methodologies to reduce departures from normality (see Foster, 1986).

[58] The definitions of high and low disclosure firms are similar to those in Gelb and Zarowin (2002).

[59] Before undertaking the analysis, I select a random sample of 100 firms. For these firms, I collect the 'total assets' from *Datastream* and I check these numbers with those reported in company annual reports. No errors are found. Additionally, I compare the 'share price' collected from *Datastream* for the same sample with those reported in *Financial Times*. I find that 99 cases are similar. This gives an indication of the reliability of *Datastream* numbers.

Earnings variables

The measure for earnings per share is *Datastream* item number 183, which is calculated by dividing number 182 'earnings for ordinary-full tax' by the number of shares outstanding. X_t, X_{t+1} and X_{t+2} are defined as earnings change deflated by share price. Both current and future earnings changes are deflated by price at the start of the return window for period *t* (see Lundholm and Myers, 2002). For example, earnings variables for the year 1996 are calculated as follows:

X_t = (EPS96-EPS95)/Price95,

X_{t+1} = (EPS97-EPS96)/Price95,

and X_{t+2} = (EPS98-EPS97)/Price95.

Where EPS is earnings per share *Datastream* item No. 183, and Price95 is the price four months after the year-end for 1995.[60]

Datastream program 900A and 900C are used to collect accounting items, while *Datastream* program 900B is used to collect daily share prices for each company.

Returns

In the present study *Datastream* program 900B is used to collect stock returns. Returns calculate the theoretical growth in the value of a share over a specific year, assuming that dividends are re-invested to purchase additional units of equity at the closing price applicable on the ex-dividend date. Return of year t is defined as the return for the 12 month period starting four months after the financial year-end of year t–1. In other words, the chosen return window extends from eight months prior to the financial year-end to four months after financial year-end. The four month lag is chosen to ensure that

[60] Stock price (P) is defined as a company's share price at the end of period (t). *Datastream* adjusts share prices for subsequent capitalisations. Consequently, past stock prices are directly comparable with current stock prices.

the information in annual report narratives have been read by the market.[61] Therefore, R_{t+1} and R_{t+2} are calculated as the buy-and-hold returns for the one-year period starting four and sixteen months after the firm's current financial year-end. Accordingly, the return windows for R_{t+1} and R_{t+2} do not overlap with the current return window. For 1996, return variables are calculated as follows:

R_t = (RI96/RI95) –1,

R_{t+1} = (RI97/RI96) –1,

and R_{t+2} = (RI98/RI97) –1

where RI is the return index for a specific year as calculated by *Datastream*.

Earnings yield

Earnings yield, EP_{t-1} is defined as earnings for the period *t*–1 divided by price four months after the financial year-end of period *t*–1. For example, EP_{t-1} for a December 1996 observation is calculated by dividing EPS for 1995 by stock price at the end of April 1996.

Asset growth

Asset growth, AG_t, is the growth rate of total book value of assets for the year t. The measure for total book value of assets is *Datastream* item No. 392. Growth of assets is calculated as the change in book value of assets in year t, divided by the book value of assets at the end of t–1. For example, asset growth in 1996 equals total assets in 1996 minus total assets 1995 divided by total assets in 1995.

[61] At an early stage, the dates under the chairmen's statements have been identified. In most cases, the annual reports are published within three months of the firm's financial year-end. One further month is assumed to be sufficient for financial analysts to read and process the information.

Disclosure dummy

Disclosure scores are converted into dummy variables. This is done separately for each of the two topic lists. Firms with a disclosure score in the top (bottom) quartile of the distribution are defined as high (low) disclosure firms.[62] The dummy variable, D, is set equal to 1 (0) for high (low) disclosure firms. Firms with disclosure scores in the second and third quartiles are deleted. Deleting firms in the middle quartiles helps to compare firms at the top end of the disclosure spectrum against firms at the bottom end. This should increase the power of the regression analysis.

6.5. Descriptive Statistics and Correlation Matrix

6.5.1. Descriptive Statistics

Table 6.2 presents descriptive statistics for two main panels with observations coming from years 1996 to 1999. The two sets differ in the definition of disclosure quality. In Panel A, disclosure scores are defined in terms of the all-inclusive topic list, while Panel B defines disclosure quality in terms of forward-looking profit topics. The number of observations is 917 and 887, respectively.

Panels A and B show that the mean annual return is 13 percent. The mean earnings change for t and t+1 are 0.6 and 0.7 percent of price. Mean stock returns and earnings change are somewhat lower in period t+2.

It is worth noting that there is a considerable variation in firm size. For example, in Panel A the market value for the first quartile is £11.4 million, while it is more than £605 million for the third quartile. Thus, unlike the AIMR-FAF database, the sample used in the present study does not focus only on the largest firms in the economy.

[62] For an identical definition of high and low disclosure see Gelb and Zarowin (2002).

Descriptive statistics for the disclosure scores indicate that there is a substantial difference between the mean disclosure scores for high and low disclosure firms. For example, the median disclosure score in Panel A for high (low) disclosure firms is 51 (9), while in Panel B the median disclosure score is 9 (1). The most important observation from Table 6.2, however, is that apart from the disclosure scores the two sets of statistics are quite similar. As a result, differences in regression parameters between Panels A and B are unlikely to come from differences other than the definition in disclosure quality.

Table 6.2. Descriptive Statistics: Pooled Data

Variable	Mean	Min.	25%	Median	75%	Max.	Obs.
Panel A: All Topics							
R_t	0.130	–0.710	–0.157	0.056	0.312	3.001	917
X_t	0.006	–0.337	–0.009	0.006	0.017	0.546	917
X_{t+1}	0.007	–0.255	–0.011	0.006	0.020	0.298	917
X_{t+2}	–0.001	–0.316	–0.019	0.003	0.021	0.254	917
R_{t+1}	0.119	–0.733	–0.182	0.048	0.326	2.647	917
R_{t+2}	0.043	–0.835	–0.235	–0.005	0.252	2.378	917
EP_{t-1}	0.049	–0.683	0.039	0.061	0.083	0.185	917
AG_t	0.151	–0.567	–0.027	0.065	0.207	2.694	917
MV (£m)	1450	0.4	11.4	80.4	605	65402	917
Disclosure = Low	8.5	0	6	9	11	15	467
Disclosure = High	55.8	33	41	51	63	168	450
Panel B: Profit Topics Only							
R_t	0.130	–0.710	–0.170	0.052	0.313	2.951	887
X_t	0.007	–0.337	–0.009	0.006	0.019	0.486	887
X_{t+1}	0.006	–0.255	–0.011	0.006	0.020	0.298	887
X_{t+2}	–0.002	–0.335	–0.020	0.003	0.020	0.254	887
R_{t+1}	0.088	–0.733	–0.226	0.010	0.285	3.076	887
R_{t+2}	0.056	–0.835	–0.233	–0.011	0.267	2.364	887
EP_{t-1}	0.048	–0.662	0.039	0.061	0.085	0.173	887
AG_t	0.157	–0.519	–0.027	0.073	0.216	2.694	887
MV (£m)	1088	0.6	14.5	68.2	378	65402	887
Disclosure = Low	0.9	0	0	1	1	2	485
Disclosure = High	10.3	7	8	9	12	31	402

Table 6.2 presents descriptive statistics for the two main panels employed in Section 6.6 using data pooled across the four-year sample period. The two sets differ in the definition of disclosure quality. In Panel A (B), disclosure quality is defined in terms of all topics (profit topics only). The earnings per share measure is *Datastream* item 183 which is item 182 'Earnings for Ordinary – Full Tax' divided by the number of shares outstanding. X_t, X_{t+1} and X_{t+2} are defined as earnings change deflated by price. Both current and future earnings changes are deflated by price at the start of the return window for period t. R_t, R_{t+1} and R_{t+2} are calculated as buy-and-hold returns (inclusive of dividends) over a 12-month period, starting four months after the end of the previous financial year. EP_{t-1} is defined as period t–1's earnings over price four months after the financial year-end of period t–1. AG_t is the growth rate of total book value of assets for period t (*Datastream* item 392). MV = market value. Disclosure = disclosure scores. Firms with a disclosure score in the top (bottom) quartile of the distribution are defined as high (low) disclosure firms. Observations with disclosure scores in the second and third quartiles are not included in the panels.

6.5.1. Correlation Matrix

Table 6.3 presents Pearson Correlations for all regression variables. Correlations above (below) the diagonal correspond to a definition of disclosure quality that includes all topics (profit topics only). P-values are given in parentheses. Correlations are estimated using data pooled across the four-year sample period. As anticipated, the correlation between the current period's return, R_t, and the current period earnings change, X_t, is significantly positive at the 5% level. On the other hand, the correlation between R_t and X_{t+1} is weaker, but still significant. But, the correlation between R_t and X_{t+2} is not significant. This might indicate that the inclusion of future earnings variables for two years is generally sufficient.

Table 6.3. Pearson Correlations: Pooled Data

	R_t	X_t	X_{t+1}	X_{t+2}	R_{t+1}	R_{t+2}	EP_{t-1}	AG_t	*MV*	*Disclo-sure*
R_t		0.236	0.078	–0.044	0.057	–0.066	0.009	0.236	0.083	0.035
		(0.001)	(0.018)	(0.187)	(0.086)	(0.047)	(0.774)	(0.001)	(0.012)	(0.288)
X_t	0.215		–0.134	–0.073	0.035	0.041	–0.485	0.071	–0.020	–0.015
	(0.001)		(0.001)	(0.028)	(0.283)	(0.213)	(0.001)	(0.032)	(0.552)	(0.645)
X_{t+1}	0.112	–0.116		–0.005	0.274	0.079	–0.212	–0.024	–0.030	–0.050
	(0.001)	(0.001)		(0.873)	(0.001)	(0.017)	(0.001)	(0.462)	(0.365)	(0.130)
X_{t+2}	–0.007	–0.041	–0.062		0.120	0.301	–0.064	–0.045	–0.017	–0.024
	(0.833)	(0.221)	(0.063)		(0.001)	(0.001)	(0.053)	(0.171)	(0.610)	(0.464)
R_{t+1}	0.069	0.023	0.290	0.088		–0.019	–0.012	–0.020	0.003	0.044
	(0.039)	(0.496)	(0.001)	(0.009)		(0.574)	(0.707)	(0.540)	(0.929)	(0.181)
R_{t+2}	–0.033	0.019	0.079	0.283	–0.001		0.036	–0.068	–0.033	–0.035
	(0.321)	(0.573)	(0.019)	(0.001)	(0.991)		(0.280)	(0.040)	(0.320)	(0.286)
EP_{t-1}	0.003	–0.483	–0.198	–0.040	–0.015	0.083		–0.024	0.001	0.004
	(0.936)	(0.001)	(0.001)	(0.239)	(0.656)	(0.014)		(0.469)	(0.965)	(0.915)
AG_t	0.171	0.083	–0.077	–0.023	–0.004	–0.040	–0.015		0.043	0.048
	(0.001)	(0.014)	(0.021)	(0.485)	(0.916)	(0.231)	(0.653)		(0.190)	(0.143)
MV	0.071	–0.021	–0.026	–0.011	–0.002	–0.037	0.005	0.022		0.480
	(0.034)	(0.523)	(0.442)	(0.751)	(0.961)	(0.267)	(0.874)	(0.514)		(0.001)
Disclo-sure	–0.070	–0.060	–0.008	–0.027	0.007	–0.031	0.124	0.019	0.264	
	(0.036)	(0.074)	(0.809)	(0.418)	(0.825)	(0.363)	(0.001)	(0.576)	(0.001)	

Table 6.3 presents Pearson Correlations for all regression variables. Correlations above (below) the diagonal correspond to a definition of disclosure quality that includes all topics (profit topics only). P-values are given in parentheses. The number of observations is 917 (above the diagonal) and 887 (below the diagonal). The earnings per share measure is *Datastream* item 183 which is item 182 'Earnings for Ordinary – Full Tax' divided by the number of shares outstanding. X_t, X_{t+1} and X_{t+2} are defined as earnings change deflated by price. Both current and future earnings changes are deflated by price at the start of the return window for period t. R_t, R_{t+1} and R_{t+2} are calculated as buy-and-hold returns (inclusive of dividends) over a 12-month period starting four months after the end of the previous financial year. EP_{t-1} is defined as period t–1's earnings over price four months after the financial year-end of period t–1. AG_t is the growth rate of total book value of assets for period t (*Datastream* item 392). MV = market value. Disclosure = disclosure scores. Firms with a disclosure score in the top (bottom) quartile of the distribution are defined as high (low) disclosure firms. Observations with disclosure scores in the second and third quartile are not included in the panels.

There is also some evidence that AG_t is not a perfect measurement error proxy. Theoretically, an errors-in-variables proxy should be highly correlated with the

measurement error but uncorrelated with the dependent variable. This is not the case for AG_t. In Table 6.3 the correlation coefficients between R_t and AG_t are 0.236 and 0.171, both significant at the 0.001 level. Nonetheless, AG_t is not deleted from the regression model in order to be consistent with Collins et al. (1994) and Gelb and Zarowin (2002).

Finally, disclosure scores are positively associated with firm size, and this correlation is significant at the 0.001 level for both definitions of disclosure quality. However, this association is much less pronounced when disclosure quality is defined in terms of profit topics, thus making the interpretation of a (correlated omitted) size effect less likely for this definition of disclosure quality.

6.6. Empirical Results

This section presents the main empirical findings for this chapter. It is divided into five subsections. Section 6.6.1 discusses the findings when the definition of disclosure quality is based an all-inclusive topic list. Section 6.6.2 discusses the results when the definition of disclosure quality is based on forward-looking profit sentences. Section 6.6.3 reports incremental R^2s. Section 6.6.4 examines the effect of other topics on the return-earnings association. Finally, Section 6.6.5 examines industry effects.

6.6.1. All-inclusive Topic List

Table 6.4 contains the regression results for the first definition of disclosure quality. Year-by-year regressions are run in addition to a pooled regression with all firm-year observations. Therefore, five sets of estimates are presented in Table 6.4. These relate to the cross-sections 1996 to 1999 and a pooled regression with observations from all four years. Pooling the observations provides a sample size that is comparable to that of

disclosure studies based on AMIR-FAF ratings. Heteroscedasticity-consistent p-values are reported in parentheses.

Table 6.4. Regression Results: All Topics.

Independent Variable	Expected Sign	Pooled	1996	1997	1998	1999
Intercept	(?)	–0.01	–0.21***	–0.07	–0.08	0.20**
		(0.736)	(0.003)	(0.104)	(0.142)	(0.044)
X_t	(+)	2.41***	3.08***	3.39***	1.51***	2.48*
		(0.001)	(0.001)	(0.001)	(0.009)	(0.067)
X_{t+1}	(+)	1.77***	2.42***	2.03***	0.22	2.49**
		(0.001)	(0.001)	(0.001)	(0.691)	(0.035)
X_{t+2}	(+)	0.28	0.74	–0.04	–0.51	0.67
		(0.321)	(0.303)	(0.911)	(0.292)	(0.629)
R_{t+1}	(–)	–0.07	–0.07	–0.03	0.15*	–0.16
		(0.203)	(0.531)	(0.757)	(0.054)	(0.387)
R_{t+2}	(–)	–0.05	0.16*	0.04	0.03	–0.39*
		(0.305)	(0.069)	(0.261)	(0.724)	(0.082)
EP_{t-1}	(+)	1.51***	3.42***	1.48***	0.85*	0.87
		(0.001)	(0.002)	(0.002)	(0.073)	(0.533)
AG_t	(–)	0.20***	0.30***	0.02	0.30**	–0.13
		(0.001)	(0.009)	(0.791)	(0.011)	(0.341)
D	(?)	0.02	0.09	0.01	0.04	–0.19
		(0.764)	(0.348)	(0.906)	(0.691)	(0.115)
$D*X_t$	(?)	1.06	1.17	0.91	3.90***	1.20
		(0.159)	(0.362)	(0.328)	(0.008)	(0.438)
$D*X_{t+1}$	(+)	–0.76	–0.06	0.16	1.20	–0.40
		(0.313)	(0.954)	(0.864)	(0.626)	(0.796)
$D*X_{t+2}$	(+)	0.15	0.38	2.48**	–0.29	2.26
		(0.804)	(0.692)	(0.036)	(0.855)	(0.200)
$D*R_{t+1}$	(?)	0.13	0.12	0.04	–0.13	–0.09
		(0.121)	(0.467)	(0.739)	(0.289)	(0.664)
$D*R_{t+2}$	(?)	–0.12	–0.30**	–0.12	–0.04	–0.28
		(0.107)	(0.018)	(0.230)	(0.777)	(0.267)
$D*EP_{t-1}$	(?)	–0.25	–0.87	1.76**	0.50	1.28
		(0.732)	(0.561)	(0.031)	(0.736)	(0.443)
$D*AG_t$	(?)	0.10	–0.01	0.01	0.25	0.37**
		(0.290)	(0.938)	(0.884)	(0.268)	(0.033)
Observations		917	243	222	216	239
Adj. R^2		0.164	0.266	0.297	0.169	0.150

Table 6.4 presents regression results for a definition of disclosure quality derived from an all-inclusive topic list. The five sets of estimates relate to the years 1996 to 1999 and a pooled regression with observations from all four years. Heteroscedasticity-consistent p-values are reported in parentheses. The dependent variable is current period return, R_t. R_t, R_{t+1} and R_{t+2} are calculated as buy-and-hold returns (inclusive of dividends) over a 12-month period starting, four months after the end of the previous financial year. X_t, X_{t+1} and X_{t+2} are defined as earnings change deflated by price. Both current and future earnings changes are deflated by price at the start of the return window for period t. EP_{t-1} is defined as period t–1's earnings over price four months after the financial year-end of period t–1. AG_t is the growth rate of total book value of assets for period t. Disclosure scores are converted into dummy variables. Firms with a disclosure score in the top (bottom) quartile of the distribution are defined as high (low) disclosure firms. The dummy variable, D, is set equal to 1 (0) for high (low) disclosure firms. Observations with disclosure scores in the second and third quartiles are not used in estimations. The significance levels (two-tail test) are: *= 10 %, ** =5 % and *** = 1 %.

As predicted, the coefficient on X_t is significantly positive at the 1% significance level in the pooled regression as well as in three of four cross-sectional regressions. There is also strong evidence of prices leading earnings by one period. This phenomenon applies to all firms and is independent of the amount of forward-looking information in annual report narratives. All five coefficients on X_{t+1} are positive, and four of these

coefficients are significant at the 5% level. The significant estimates for this coefficient range from 1.77 to 2.49. There is no evidence that prices lead future earnings changes by more than one period. The coefficients on X_{t+2} are insignificant in all cases.

Table 6.4 shows that the coefficients on EP_{t-1} are positive as anticipated. The results also show that the coefficients on AG_t have generally the wrong sign. This suggests that AG_t might not be a satisfactory measurement error proxy.

The coefficients on the future stock return variables are expected to be negative. Many negative coefficients on future stock returns could demonstrate that realised future earnings contain a measurement error that future returns remove (Collins et al., 1994). However, the coefficients on future stock returns are mixed and largely insignificant. In the pooled regression, they have the right sign but are insignificant.

The coefficients of primary interest in the current study are the coefficients on D^*X_{t+1} and D^*X_{t+2}. A positive coefficient is hypothesised. Unfortunately, in the four cross-sections the coefficients on the two interaction terms are positive and negative, but – with one exception – not significant at conventional levels. Even in the pooled regression, the incremental future ERCs for high disclosure firms are insignificant. The coefficients on D^*X_{t+1} and D^*X_{t+2} are –0.76 and 0.15 and the corresponding p-values are 0.313 and 0.804. Thus, with an all-inclusive definition of disclosure quality, there is no evidence that the stock prices of high disclosure firms have significantly greater forecasting power for future earnings changes than those of low disclosure firms. These results are not consistent with hypothesis 1, that forward-looking information in annual report narratives enables the market to better anticipate future earnings changes. On the other hand, these results are consistent with the findings in

Schleicher (1996), Gelb and Zarowin (2002) and Haw et al. (2002). These studies are also unable to find the hypothesised association between annual report disclosures and prices leading earnings.

6.6.2. Profit Related Disclosure Topics

This section presents regression results for the second definition of disclosure quality. This definition focuses only on forward-looking sentences in relation to margins, trading results and profits.

Table 6.5 presents the findings in relation to this definition of disclosure quality. The table shows that the coefficients on $D*X_{t+1}$ and $D*X_{t+2}$ in the four individual years have mixed signs and are not significant at any conventional levels. For the pooled regression, however, the coefficient on $D*X_{t+1}$ is 1.69 with a p-value of 0.042. This coefficient implies that the market is able to anticipate more than twice as much of the earnings change for period *t*+1 as a result of profit-related forecasts. The coefficient on $D*X_{t+2}$ is positive, but not significant. This is, however, in line with my prior expectation for this definition of disclosure quality. Finance directors do not like to make profit predictions beyond the immediate future. In many cases, the 'forward-looking' sentences relate only to the first quarter of the new financial year. Making trading statements only about the immediate future then reduces the risk of being held accountable for inaccurate forecasts.

Overall, it appears that the use of disclosure scores based on *all forward-looking information* introduces too much noise into the scoring sheet. On the other hand, disclosure scores based on profit topics only yield results consistent with hypothesis 2. The results in Table 6.5 clearly show that profit-related forecasts are value-relevant to

the market. In the following section, I examine whether similar results are obtained when focusing on R^2s.

Table 6.5. Regression Results: Profit Topics Only

Independent Variable	Expected Sign	Pooled	1996	1997	1998	1999
Intercept	(?)	0.06*	–0.14***	–0.07	–0.10*	0.31***
		(0.062)	(0.009)	(0.120)	(0.080)	(0.005)
X_t	(+)	2.35***	2.85***	3.38***	2.48***	1.97
		(0.001)	(0.001)	(0.001)	(0.002)	(0.255)
X_{t+1}	(+)	1.52***	2.13***	1.97***	1.67***	3.65
		(0.001)	(0.002)	(0.001)	(0.003)	(0.108)
X_{t+2}	(+)	0.27	1.34*	0.52*	0.33	1.27
		(0.390)	(0.086)	(0.053)	(0.339)	(0.288)
R_{t+1}	(–)	0.01	–0.12	0.04	0.09	–0.20
		(0.851)	(0.333)	(0.622)	(0.270)	(0.347)
R_{t+2}	(–)	–0.05	0.08	0.01	–0.03	–0.74***
		(0.263)	(0.343)	(0.657)	(0.735)	(0.006)
EP_{t-1}	(+)	1.35***	2.83***	2.03***	1.85***	1.85*
		(0.001)	(0.002)	(0.001)	(0.001)	(0.087)
AG_t	(–)	0.18**	0.31	–0.04	0.15	–0.18
		(0.013)	(0.113)	(0.484)	(0.314)	(0.227)
D	(?)	–0.09*	–0.03	0.03	0.02	–0.41***
		(0.095)	(0.644)	(0.728)	(0.822)	(0.006)
$D*X_t$	(?)	0.54	1.09	0.88	1.25	2.19
		(0.485)	(0.263)	(0.416)	(0.451)	(0.286)
$D*X_{t+1}$	(+)	1.69**	0.71	–0.77	0.36	2.21
		(0.042)	(0.446)	(0.503)	(0.802)	(0.420)
$D*X_{t+2}$	(+)	0.63	0.90	–1.52	–0.55	2.73*
		(0.249)	(0.331)	(0.154)	(0.686)	(0.087)
$D*R_{t+1}$	(?)	–0.03	0.12	0.14	–0.03	–0.02
		(0.687)	(0.434)	(0.244)	(0.784)	(0.936)
$D*R_{t+2}$	(?)	–0.11*	–0.19	–0.09	–0.01	–0.03
		(0.085)	(0.116)	(0.291)	(0.980)	(0.911)
$D*EP_{t-1}$	(?)	–0.23	0.15	–0.33	–1.30	2.21
		(0.745)	(0.901)	(0.707)	(0.349)	(0.221)
$D*AG_t$	(?)	0.06	–0.15	0.11	0.35	0.37**
		(0.570)	(0.466)	(0.243)	(0.154)	(0.037)
Observations		887	212	230	207	235
Adj. R^2		0.131	0.273	0.272	0.142	0.178

Table 6.5 presents regression results for a definition of disclosure quality based on forward-looking profit sentences. The five sets of estimates relate to the years 1996 to 1999 and a pooled regression with observations from all four years. Heteroscedasticity-consistent p-values are reported in parentheses. The dependent variable is current period return, R_t. R_t, R_{t+1} and R_{t+2} are calculated as buy-and-hold returns (inclusive of dividends) over a 12-month period, starting four months after the end of the previous financial year. X_t, X_{t+1} and X_{t+2} are defined as earnings change deflated by price. Both current and future earnings changes are deflated by price at the start of the return window for period t. EP_{t-1} is defined as period t–1's earnings over price four months after the financial year-end of period t–1. AG_t is the growth rate of total book value of assets for period t. Disclosure scores are converted into dummy variables. Firms with a disclosure score in the top (bottom) quartile of the distribution are defined as high (low) disclosure firms. The dummy variable, D, is set equal to 1 (0) for high (low) disclosure firms. Observations with disclosure scores in the second and third quartiles are not used in estimations. The significance levels (two-tail test) are: *= 10 %, ** =5 % and *** = 1 %.

6.6.3. The Incremental R^2s

This section provides further evidence that forward-looking profit disclosures are associated with prices leading earnings. Separate regressions are run for high and low disclosure firms and for two regression models. The first model simply regresses current

stock returns on current earnings. Equation (1) is usually referred to as the simple return-earnings regression:

$$R_t = b_0 + b_1 X_t \quad (1)$$

The second model is based on the ideas in Collins et al. (1994). It regresses current stock returns on current and future earnings variables. Equation (2) is given by:

$$R_t = b_0 + b_1 X_t + \sum_{k=1}^{2} b_{k+1} X_{t+k} + \sum_{k=1}^{2} b_{k+3} R_{t+k} + b_6 EP_{t-1} + b_7 AG_t \quad (2)$$

The difference in the explanatory power between the two regressions can be ascribed to 'prices leading earnings'. This difference is termed 'incremental R^2'. The incremental R^2 is estimated for high and low disclosure firms using the following three steps:

(a) model (1) is used to compute the simple return-earnings R^2,

(b) model (2) is used to calculate the full model R^2, and

(c) the incremental R^2 is determined by subtracting the simple return-earnings R^2 from the full model R^2.

The main prediction is that if firms provide more forward-looking information in their annual report discussion section, then one would expect more future earnings news to be reflected in current returns. Hence, the incremental R^2 should be high for firms with high levels of disclosure. On the other hand, if firms disclose little forward-looking information in their annual reports, then one would expect less future earnings news to be reflected in current returns. Hence, the incremental R^2 should be smaller.

Panel A of Table 6.6 shows R^2s for disclosure based on an all-inclusive topics list. It indicates that the inclusion of future earnings variables in the regression model increases the explanatory power considerably. However, the resulting incremental R^2s

are very similar for the two groups of high and low disclosure firm-years, suggesting that an all-inclusive topic list is unable to discriminate between relevant and irrelevant disclosures.

Table 6.6. The Incremental R^2

	High Disclosure	Low Disclosure	High – Low
Panel A: All Topics			
The simple return-earnings model R^2	0.054	0.065	
The full model R^2	0.153	0.174	
Incremental R^2	0.099	0.109	–0.011
Panel B: Profit Topics Only			
The simple return-earnings model R^2	0.058	0.038	
The full model R^2	0.190	0.095	
Incremental R^2	0.132	0.058	0.074

The full model is:

$$R_t = b_0 + b_1 X_t + \sum_{k=1}^{2} b_{k+1} X_{t+k} + \sum_{k=1}^{2} b_{k+3} R_{t+k} + b_6 EP_{t-1} + b_7 AG_t$$

The simple return-earnings model is:

$$R_t = b_0 + b_1 X_t .$$

Table 6.6 presents (adjusted) R^2s from pooled regressions. Regressions are run separately for high and low disclosure firm-years from the period 1996–99. Incremental R^2s are calculated as the difference in the (adjusted) R^2 between a simple return-earnings regression and an augmented regression that includes proxies for changes in expected future earnings. In both regressions the dependent variable is current period return, R_t. R_t, R_{t+1} and R_{t+2} are calculated as buy-and-hold returns (inclusive of dividends) over a 12-month period starting four months after the end of the previous financial year. X_t, X_{t+1} and X_{t+2} are defined as earnings change deflated by price. Both current and future earnings changes are deflated by price at the start of the return window for period t. EP_{t-1} is defined as period t–1's earnings over price four months after the financial year-end of period t–1. AG_t is the growth rate of total book value of assets for period t. Firms with a disclosure score in the top (bottom) quartile of the distribution are defined as high (low) disclosure firms. Observations with disclosure scores in the second and third quartiles are not included in any regression.

In Panel B results are shown for profit topics only. Here, the incremental R^2 is more than twice as large for high disclosure firms than for low disclosure firms (0.132 versus 0.058). This indicates that the future earnings variables are relatively more important for firms with profit-related forecasts. Therefore, these results provide additional support for the main findings derived from coefficients estimates. They show that only enhanced disclosure about forward-looking profit topics is associated with stock prices that are more informative about future earnings changes.

6.6.4. Other Classes of Information

Having established a positive relation between forward-looking profit statements and 'prices leading earnings', it is natural to ask whether a similar association exists for other items from the profit and loss account. This section tries to answer this question. Here, two major line items in profit, turnover and costs, are examined. Accordingly, two further sets of disclosure scores are calculated via *Nudist*.

The first score is related to the number of forward-looking sentences relating to turnover. The annual report discussion section is text-searched for sentences that include a turnover topic such as 'turnover', 'revenue' and 'sales'. The second score is related to cost topics. The topics include 'cost', 'charge', 'expense', 'overhead' and 'saving'. As always, I only focus on forecasts.

Because the number of words and scores is relatively small, high and low disclosure scores are defined as the top and the bottom 16.67 percent of the distribution of disclosure scores. This reduces the number of observations in the pooled regression, but increases the spread of disclosure scores between high and low disclosure firm-years. For the purpose of comparison, the regression coefficients for the profit definition of disclosure quality are also re-estimated. The results for the profit, turnover and cost topic lists are presented in Table 6.7.[63]

For the profit-related scores, the coefficient estimate on D^*X_{t+1} is now larger and more significant than the one in Table 6.5, with the coefficient on D^*X_{t+1} increasing from 1.69 to 1.90. The level of significance increases from 0.042 to 0.037. This is what one expects if one increases the spread between high and low disclosure firm-years.

[63] After deleting 66.67 percent of observations in the middle range, the difference in the median disclosure score between high and low disclosure firm-years is 9, 5 and 6 for the profit, turnover and cost definition of disclosure quality.

Table 6.7. Other Classes of Information

Independent Variable	Expected Sign	Profit	Turnover	Costs
Intercept	(?)	0.02	0.02	–0.06*
		(0.502)	(0.536)	(0.067)
X_t	(+)	2.49***	2.13***	2.67***
		(0.001)	(0.001)	(0.001)
X_{t+1}	(+)	1.67***	2.16***	2.58***
		(0.001)	(0.001)	(0.001)
X_{t+2}	(+)	0.55*	0.10	–0.03
		(0.091)	(0.795)	(0.941)
R_{t+1}	(–)	0.01	–0.03	–0.04
		(0.815)	(0.639)	(0.539)
R_{t+2}	(–)	–0.04	–0.17***	–0.06
		(0.451)	(0.004)	(0.200)
EP_{t-1}	(+)	1.58***	1.73***	2.10***
		(0.001)	(0.001)	(0.001))
AG_t	(–)	0.21**	0.28***	0.21**
		(0.025)	(0.001)	(0.023)
D	(?)	–0.09*	0.03	0.03
		(0.088)	(0.675)	(0.618)
$D*X_t$	(?)	0.90	0.46	0.88
		(0.339)	(0.581)	(0.371)
$D*X_{t+1}$	(+)	1.90**	–0.67	–1.49*
		(0.037)	(0.386)	(0.085)
$D*X_{t+2}$	(+)	0.44	–0.38	1.62**
		(0.484)	(0.571)	(0.014)
$D*R_{t+1}$	(?)	–0.06	0.01	–0.06
		(0.504)	(0.873)	(0.462)
$D*R_{t+2}$	(?)	–0.13*	0.11	–0.13
		(0.057)	(0.197)	(0.112)
$D*EP_{t-1}$	(?)	0.28	–0.82	–0.97
		(0.690)	(0.297)	(0.345)
$D*AG_t$	(?)	–0.01	–0.09	–0.09
		(0.996)	(0.481)	(0.394)
Observations		676	662	519
Adj. R^2		0.160	0.132	0.211

Table 6.7 presents regression results for two other definitions of disclosure quality. Column 4 (5) focuses on forward-looking sentences in relation to turnover (costs). The estimates relate to pooled regressions with observations from the years 1996 to 1999. Heteroscedasticity-consistent p-values are reported in parentheses. The dependent variable is current period return, R_t. R_t, R_{t+1} and R_{t+2} are calculated as buy-and-hold returns (inclusive of dividends) over a 12-month period starting four months after the end of the previous financial year. X_t, X_{t+1} and X_{t+2} are defined as earnings change deflated by price. Both current and future earnings changes are deflated by price at the start of the return window for period t. EP_{t-1} is defined as period t–1's earnings over price four months after the financial year-end of period t–1. AG_t is the growth rate of total book value of assets for period t. Disclosure scores are converted into dummy variables. Firms with a disclosure score in the top (bottom) 16.67 percent of the distribution are defined as high (low) disclosure firms. The dummy variable, D, is set equal to 1 (0) for high (low) disclosure firms. Observations with disclosure scores in the middle range are not used in estimations. Column 3 re-estimates the regression coefficients for the profit definition of disclosure quality with 66.67 percent of the observations in the middle ranges deleted. The significance levels (two-tail test) are: *= 10 %, ** =5 %and *** = 1 %.

Somewhat surprisingly, there is no evidence that predictions about turnover help the market to anticipate future earnings changes more accurately. The coefficients on $D*X_{t+1}$ and $D*X_{t+2}$ are both insignificant.

For the cost-related definition of disclosure quality, the coefficients are negative at –1.49 for the period t+1and positive at 1.62 for the period t+2. The corresponding p-

values are 0.085 and 0.014. Such estimates are consistent with a scenario where the announcement of a cost saving program is perceived by the market as a positive NPV project. This results in positive returns in the announcement period *t*. While cash outlays and provisions reduce period *t*+1 earnings, the benefits mainly accrue in the following financial year.[64]

6.6.5. Industry Control

A potential problem with the estimations so far is that they ignore cross-sectional differences in the predictability and timeliness of earnings. Firms with more predictable and/or less timely earnings should exhibit a higher association between current returns and future earnings. If these two factors are correlated with disclosure quality scores, then the results might be driven by differences in the fundamental lead-lag relation between returns and earnings rather than increased voluntary disclosure. Therefore, it is important to control for the differences in earnings timeliness. To do that, all firms are classified into nine broadly defined industries sectors. I then re-estimate the coefficients holding the industry classification constant. It is argued that grouping firms by industry helps to isolate the effect of disclosure quality on 'prices leading earnings' because it eliminates inter-industry differences in accounting and real business factors (Gelb and Zarowin, 2002).

Table 6.8 reports the regression results for three different degrees of industry control. Column 3 evaluates a firm's disclosure quality relative to its industry peers by ranking disclosure scores separately for each industry sector. Column 4 allows the intercept and the current ERC to vary across industries. Finally, Column 5 evaluates a firm's

[64] Note that the returns in period *t* are calculated over a 12-month period that extends 4 months into the financial year *t*+1. Thus, the interpretation of the 'cost coefficients' is perfectly consistent with the conservatism principle that requires firms to expense investments with uncertain future benefits immediately.

Table 6.8. Specification Test: Industry Control

Independent Variable	Expected Sign	Industry Control: Disclosure Scores	Industry Control: Disclosure Scores Current Earnings	Industry Control: Disclosure Scores Current Earnings Future Earnings
Intercept	(?)	0.04 (0.202)	0.05 (NA)	0.01 (NA)
X_t	(+)	2.63*** (0.001)	2.44 (NA)	2.44 (NA)
X_{t+1}	(+)	1.40*** (0.001)	1.57*** (0.001)	1.77 (NA)
X_{t+2}	(+)	0.35 (0.287)	0.45 (0.135)	–0.12 (NA)
R_{t+1}	(–)	–0.02 (0.757)	–0.06 (0.403)	–0.05 (NA)
R_{t+2}	(–)	–0.05 (0.319)	–0.08 (0.120)	–0.04 (NA)
EP_{t-1}	(+)	1.39*** (0.001)	1.69*** (0.001)	2.73 (NA)
AG_t	(–)	0.20** (0.018)	0.14* (0.085)	0.11 (NA)
D	(?)	–0.10** (0.034)	–0.14** (0.015)	–0.12** (0.011)
$D*X_t$	(?)	1.35** (0.042)	1.38** (0.049)	1.30** (0.047)
$D*X_{t+1}$	(+)	1.83** (0.011)	2.00*** (0.007)	2.18*** (0.001)
$D*X_{t+2}$	(+)	0.53 (0.323)	0.63 (0.255)	0.48 (0.383)
$D*R_{t+1}$	(?)	0.01 (0.857)	–0.02 (0.819)	0.03 (0.690)
$D*R_{t+2}$	(?)	–0.07 (0.302)	–0.05 (0.417)	–0.01 (0.840)
$D*EP_{t-1}$	(?)	0.56 (0.333)	1.15 (0.140)	0.53 (0.357)
$D*AG_t$	(?)	0.01 (0.949)	0.02 (0.809)	0.03 (0.760)
Adj. R^2		0.161	0.237	0.271
Observations		874	874	874

Table 6.8 presents a specification test. Regression results are reported for three different degrees of industry control. Column 3 assesses a firm's disclosure quality relative to its industry peers by ranking disclosure scores separately for each industry sector. Column 4 also allows the intercept and the current ERC to vary across industries. In Column 5 a firm's disclosure quality is evaluated relative to its industry and both current and future ERCs are allowed to vary. The estimates relate to a pooled regression with observations from the years 1996 to 1999. Where estimates are allowed to vary, average coefficients are reported. Heteroscedasticity-consistent p-values are reported in parentheses. In all three columns the disclosure quality is defined in terms of forward-looking profit sentences. The dependent variable is current period return, R_t. R_t, R_{t+1} and R_{t+2} are calculated as buy-and-hold returns (inclusive of dividends) over a 12-month period starting four months after the end of the previous financial year. X_t, X_{t+1} and X_{t+2} are defined as earnings change deflated by price. Both current and future earnings changes are deflated by price at the start of the return window for period t. EP_{t-1} is defined as period t–1's earnings over price four months after the financial year-end of period t–1. AG_t is the growth rate of total book value of assets for period t. Disclosure quality is calculated by adding up the number of forward-looking profit sentences in annual report narratives. Disclosure scores are converted into dummy variables. Firms with a disclosure score in the top (bottom) quartile of the distribution are defined as high (low) disclosure firms. The dummy variable, D, is set equal to 1 (0) for high (low) disclosure firms. Observations with disclosure scores in the second and third quartile are not used in estimations. The significance levels (two-tail test) are: *= 10 %, ** =5 % and *** = 1 %.

disclosure score relative to its industry and both current and future ERCs are allowed to vary. In all three columns, the disclosure quality is defined in terms of forward-looking information related to profit topics.

The regression results in Table 6.8 show that the earlier findings on forward-looking profit disclosures become even stronger after controlling for industry effects. All three coefficients on $D^{*}X_{t+1}$ are now larger and more significant than the corresponding estimates in Table 6.5. Furthermore, the period *t*+1 incremental ERC increases and becomes more significant when one moves from Column 3 to Column 5. If anything, the results in Table 6.8 suggest that the failure to control for accounting and business factors has weakened the earlier results.

6.7. Summary

A major contribution of the thesis is the development of a scoring methodology that partially automates the generation of disclosure scores and thus allows the researcher to produce disclosure scores for large samples of firms. This chapter uses these disclosure scores to examine the information content of annual report narratives for future earnings. Following Lundholm and Myers (2002) and Gelb and Zarowin (2002) I examine the association between the disclosure of forward-looking information in annual report narratives and prices leading earnings. The first hypothesis in this chapter predicts that there is a positive association between forward-looking disclosures and prices leading earnings. However, the empirical findings are not in line with this association. The most likely explanation is that annual report narratives discuss a wide range of topics. Most of these topics are at best weakly correlated with future earnings changes.

The second hypothesis predicts that there is a positive association between forward-looking profit sentences and prices leading earnings. The findings in Section 6.6.2 are in line with that prediction. These results suggest that the ability of the market to anticipate one-year-ahead future earnings changes is positively related to the quality of disclosure

when disclosure is defined in terms of forward-looking profit sentences. These results are further supported by the analysis of incremental R^2s.

The association between other items from the profit and loss account and prices leading earnings is also examined. In this analysis, the definition of disclosure quality is based on two other disclosure indices. These indices contain topics related to turnover and costs. For the turnover definition, there is no evidence that predictions about sales help the market to forecast future earnings changes more accurately. When I define disclosure quality in terms of cost topics, I then obtain significant coefficients which are consistent with the idea that the market perceives cost saving programs as positive NPV projects.

Finally, firms are assigned to industry sectors and I re-estimate the coefficients holding the industry classification constant. The findings indicate that the results on forward-looking profit disclosures become even stronger after controlling for industry effects.

Overall, the results of the cross-sectional analysis in this chapter are important because they suggest that the automated text search approach developed in Chapters 4 and 5 has a potential to identify value-relevant disclosures.

Chapter 7: Disclosure Quality and Prices Leading Earnings: A Time-Series Analysis

7.1. Overview

The cross-sectional results in Chapter 6 suggest that the ability of the market to anticipate one-year-ahead future earnings changes is positively related to the quality of disclosure when disclosure is defined in terms of forward-looking profit statements. To complement the cross-sectional results in Chapter 6, the present chapter relates changes in voluntary corporate disclosures to changes in the extent to which returns anticipate future earnings changes. If the level of forward-looking disclosure affects the degree to which future earnings are reflected in current returns, then an increase (decrease) in the level of disclosure should result in an increase (decrease) in the extent to which stock returns anticipate future earnings changes. This proposition is tested by examining changes in the level of disclosure over the sample period 1996–1999.

The analysis in this chapter follows Lundholm and Myers (2002) in examining the effect of changes in disclosure quality on the relation between returns and future earnings. The sample is sorted into two groups, disclosure increasers and disclosure non-increasers. For these two groups, I estimate the incremental R^2 in the first and the last year that a firm appears in the dataset. The incremental R^2 is defined as the difference between the R^2 from the full regression model with future earnings variables included as independent variables and the R^2 from the simple return-earnings regression with current earnings as the only independent variable.

The empirical results for the sample period 1996–1999 are not in line with the hypothesised prediction regarding the association between changes in disclosure quality and the relation between returns and future earnings.

In an attempt to explain these unexpected results, several modifications of the original research design are considered. These include: (a) extending the time-span between the first and the last year in the analysis, (b) choosing a new sample period, (c) redefining disclosure changes according to their ranks in the first and the last year of the analysis, and (d) re-categorising firms according to the change in the number of 'strong' sentences as defined through labour-intensive reading.

The final modification provides evidence that changes in a firm's forward-looking disclosure is positively related to changes in the amount of future earnings news reflected in current earnings. Overall, the findings in this chapter suggest that automated text-searches might not be able to pick up subtle differences in disclosure quality that appear important in a time-series framework where the amount of information changes is relatively small.

The remainder of this chapter is organised as follows. Section 7.2 provides a review of the time-series analysis in Lundholm and Myers (2002). The regression models and the hypothesis are discussed in Section 7.3. Section 7.4 describes the data. The empirical results for the period 1996–1999 are presented in Section 7.5. Sections 7.6 to 7.9 present the findings in relation to the modifications of the original research design. Section 7.10 summarises.

7.2. Lundholm and Myers (2002)

There are relatively few studies that examine the effect of changes in disclosure scores over time. One such study is Healy, Hutton and Palepu (1999). Healy et al. (1999) show that firms that increase their disclosure levels over time experience lower bid-ask spreads. Another study that focuses on time-series analysis is Leuz and Verrecchia

(2000). Leuz and Verrecchia (2000) also find that firms committing to a higher disclosure level experience lower bid-ask spreads and higher trading volumes.

This chapter follows the time-series analysis in Lundholm and Myers (2002). They examine how changes in disclosure quality affect the amount of future earnings changes being reflected in current returns. More specifically, they test the hypothesis that changes in disclosure are positively related to changes in the proportion of the variation in current returns that can be explained by variations in future earnings.

To test this hypothesis, Lundholm and Myers (2002) select the first and the last year that a firm is in their AIMR disclosure dataset, so as to measure the effect of changes in corporate disclosure over the longest possible time span. They then sort these firms into disclosure increasers (303 firms) and disclosure decreasers (290 firms). For these two groups, they estimate the contribution of future earnings for returns over and above the contribution of current earnings in the first and the last year that a firm exists in their dataset. They define this contribution as the difference between the R^2 from the full regression model with future earnings variables included as regressors (see Section 6:6) and the R^2 from the simple return-earnings regression model with current earnings as the only regressor. The difference in R^2 between the two models indicates the importance of future earnings for current returns. Lundholm and Myers (2002) term the difference in R^2 between the two models as the 'Incremental R^2'.

The authors argue that if firms' disclosures bring future earnings news forward into current stock returns, then the incremental R^2 should increase for the disclosure increasers. On the other hand, forecasting the direction of the incremental R^2 for the disclosure decreasers is more difficult. They argue that, despite the fact that the disclosure practice of this group has declined compared with their industry peers, the

incremental R^2 could increase for this group as well. This could arise due to external effects such as an economy-wide increase in corporate disclosure. To control for an economy-wide change in corporate disclosure, the authors predict that the change in the incremental R^2 for the disclosure increasers would be greater than the change in the incremental R^2 for the disclosure decreasers.

The regression results in Lundholm and Myers (2002) indicate that the ability of future earnings news to explain current returns increases for the disclosure increasers. Using the industry-adjusted data, the authors find that the incremental R^2 increases from 0.0665 in the first year to 0.1906 in the last year. When they use raw data, they find the incremental R^2 increases from 0.0908 in the first year to 0.2952 in the last year. For the disclosure decreasers, the change in the incremental R^2 is negative for the industry-adjusted data, decreasing from 0.1593 to 0.1149 and approximately zero for the raw data, increasing from 0.1299 to 0.1314. Overall, the authors find evidence that changes in firm disclosure are positively related to changes in the importance of future earnings news for current returns.

The present study generally follows Lundholm and Myers (2002) in examining the effect of changes in disclosure quality. However, my study differs from Lundholm and Myers (2002) in that it focuses on forward-looking information published in the annual report discussion section. Also, it emphasises only one particular class of forward-looking information. This class relates to forecasts of profit related topics. In Lundholm and Myers (2002), AIMR-FAF ratings are used. These ratings are based on evaluating three media of corporate disclosures. These are annual reports, quarterly reports and investor relations (see Section 2.4.2).

7.3. Research Design

In order to test the effect of changes in disclosure quality on the relation between returns and future earnings changes, I focus on the change in disclosure between 1996 and 1999. These years are the first and the last year in the cross-sectional analysis in Chapter 6.

Based on the change in disclosure scores, firms are sorted into three main groups. These groups are disclosure increasers, disclosure decreasers and firms with no change in their disclosure scores over time. The three groups are defined as follows: (1) disclosure increasers refer to firms that increase their disclosure scores between 1996 and 1999, (2) disclosure decreasers refer to firms that decrease their disclosure scores between 1996 and 1999, and (3) firms with no change in their disclosure maintain their disclosure scores between 1996 and 1999.

In order to measure how changes in corporate disclosure change the relation between current stock returns and future earnings, two regression models are used. The first model simply regresses current stock return on current earnings. Equation (1) is usually referred to as the simple return-earnings regression:

$$R_t = b_0 + b_1 X_t \quad (1)$$

The second model regresses current stock return on current and future earnings variables.

$$R_t = b_0 + b_1 X_t + \sum_{k=1}^{2} b_{k+1} X_{t+k} + \sum_{k=1}^{2} b_{k+3} R_{t+k} + b_6 EP_{t-1} + b_7 AG_t \quad (2)$$

Equation (2) is similar to the main regression model in Chapter 6 but the cross-sectional analysis in Chapter 6 allows the coefficients to vary with disclosure quality.

Following Lundholm and Myers (2002), I estimate the contribution of the future earnings variables X_{t+1}, X_{t+2}, R_{t+1}, R_{t+2}, EP_{t-1} and AG_t to the full model R^2 in 1996 and 1999. In order to calculate the incremental R^2, three steps are followed for each disclosure group and year:

(a) model (1) is used to compute the simple return-earnings R^2,
(b) model (2) is utilised to calculate the full model R^2, and
(c) the incremental R^2 is determined by subtracting the simple return-earnings R^2 from the full model R^2.

The main prediction is that if firms increase their disclosure over time, one would expect more future earnings news to be reflected in current returns. Hence, the incremental R^2 should increase. On the other hand, if firms decrease their disclosure over time, one would expect relatively less future earnings news to be reflected in current returns.

Based on the voluntary information in the annual report discussion section alone, one could expect an increase in the incremental R^2 for disclosure increasers and a decrease in the incremental R^2 for disclosure decreasers. However, this ignores the fact that the amount of information about firms might change through time as a result of other trends, for example through an increase of mandatory disclosure requirements over time. To account for such a trend, the hypothesis in this chapter only predicts that the change in the incremental R^2 of disclosure increasers between 1996 and 1999 is greater than the change in the incremental R^2 for disclosure decreasers during the same period.

7.4. Data

7.4.1. Sample Selection

The years 1996 and 1999 are the first and the last year in the cross-sectional analysis in Chapter 6. For that analysis, 800 firms are randomly selected per year. However, to increase the total number of firms available for the time-series test, I select all non-financial firms that have an annual report in 1996 and in 1999. The time span between the first and the last year is three years.

The sample selection procedure is described in Table 7.1. The total number of non-financial firms on Dialog for 1996 and 1999 is 1122 and 1289, respectively. Disclosure change is defined as the difference between disclosure scores in 1999 and 1996. Hence, firms in 1996 are matched with those in 1999. To be included in the sample, firms must have an annual report in both 1996 and 1999. Firms with only one annual report in either of these years are excluded from the sample. The resulting sample consists of 784 non-financial firms.

Five firms are excluded from the sample because of the unavailability of *Datastream* codes. The remaining sample comprises 779 non-financial firms. These firms are scored via *Nudist* and disclosure scores are exported into a spreadsheet. Finally, these scores are combined with the firm's earnings per share and return data using SAS. The definition of accounting and return variables is the same as for the cross-sectional analysis.

Further reductions arise mainly due to changing year-ends (107 firms) and missing observations for any of the regression variables (209 firms). Consistent with Chapter 6, outliers are defined as the top and the bottom 1% of observations for the distribution of

any of the regression variables. The final sample available to carry out the analysis thus comprises 367 firms.

It is important to note that excluding a firm in one year for any of the above reasons leads to the exclusion of the same firm from the other year. This is because matched observations are needed to undertake the analysis. For example, the total number of firms with missing values in 1996 is 192 firms. These firms are automatically excluded from 1999 as well. In 1999 there are a further 17 firms with missing observations. As a result, the reported number of missing observations in Table 7.1 equals 209 firms.

Table 7.1. Sample Selection Criteria: Sample Period 1996–1999

	1996	1999
Total number of firms on Dialog	1594	1892
(–) Financial firms	(472)	(603)
= Non-financial firms	1122	1289
(–) Firms with only one annual report in one year in each sample period	(338)	(505)
= Matched firms within sample period	784	784
(–) Firms with no *Datastream* codes	(5)	(5)
= Matched firms	**779**	**779**
(–) Firms that changed their year-ends	(107)	(107)
(–) Firms with missing observations	(209)	(209)
= Number of firms before deleting outliers	463	463
(–) Outliers	(96)	(96)
= Firms to be used in the regression analysis	**367**	**367**

Table 7.1 presents the sample selection criteria. It starts with the total number of firms on Dialog. Financial firms are then excluded. Further observations are subsequently deleted due to the unavailability of annual reports, the unavailability of *Datastream* information, changing financial year-ends and outliers. Excluding a firm in one year for any of the above reasons leads automatically to the exclusion of the same firm from the other year.

7.4.2. Descriptive Statistics

As indicated earlier, *Nudist* is used to generate disclosure scores for the selected firms in the sample. The total number of firms available for the time-series analysis is 367.

Initially, both sets of topics are used to score annual report narratives. The first set is based on an all-inclusive topic list, while the second list is based on profit topics only. For each year, running command files in *Nudist* leads to the generation of disclosure scores for each firm. Then the change in disclosure scores is calculated for each firm. Finally, firms are divided into three sub-categories. The first category includes firms that increase their disclosure levels between 1996 and 1999. The second group represents the firms that decrease their disclosure levels between 1996 and 1999. The final category includes firms that maintain their disclosure levels between 1996 and 1999.

Table 7.2 shows the total number of firms in each sub-category. It consists of two panels. Panel A is related to the disclosure definition based on an all-inclusive topic list, while Panel B is related to the disclosure definition based on profit topics only.

Panel A of Table 7.2 shows that with an all-inclusive definition of disclosure quality, 90.2% of firms increase their disclosure scores from 1996 to 1999. Only a small number of firms maintain or decrease their disclosure levels between 1996 and 1999. This means that there is generally an improvement in disclosure levels over time when disclosure is measured using the all-inclusive topic list.

Panel B of Table 7.2 shows that with a profit-based definition of disclosure quality, the number of firms that increase their disclosure scores between 1996 and 1999 is roughly equal to the total number of firms that do not increase their disclosure levels. From now on, the analysis only focuses on the definition of disclosure quality based on profit topics only. The reasons are twofold: First, the number of observations in the increasers and non-increasers groups is roughly equal for this definition of disclosure quality. Secondly, the analysis in Chapter 6 provides evidence that a positive relation between

prices leading earnings and disclosure quality only exists when disclosure quality is measured in terms of profit topics.

Table 7.2. Disclosure Changes between 1996 and 1999

	Firms	%
Panel A: Disclosure based on an all-inclusive topic list		
Increasers	331	90.2
Decreasers	31	8.4
Firms with no change in their disclosure scores	5	1.4
Total	**367**	**100**
Panel B: Disclosure based on profit topics only		
Increasers	188	51.2
Decreasers	136	37.1
Firms with no change in their disclosure scores	43	11.7
Total	**367**	**100**

Table 7.2 classifies firms into disclosure increasers, disclosure decreasers and firms with no change in disclosure scores. This classification is given for two definitions of disclosure quality. The first definition is based on an all-inclusive topic list, while the second definition is based on the profit topic list only. Column 2 shows the number of firms in each group. Column 3 shows the percentage of firms in each group.

Table 7.3 displays summary statistics for the distribution of disclosure scores based on the profit definition of disclosure quality. Panel A shows that disclosure scores in 1996 range from a minimum value of 0 to a maximum value of 31. The median value is 4 and the average value is 4.56.

Panel B shows that the distribution is very similar for 1999. Median and average values are now 4 and 5.39, respectively. Of course, in a time-series analysis, the more important statistic relates to changes in disclosure quality over time. Descriptive statistics on changes are reported in Table 7.4.

Table 7.4 displays descriptive statistics on the distribution of changes in disclosure quality for each disclosure group. For the increasers group, increases in disclosure scores range from a minimum of 1 up to a maximum of 18. The average increase equals

4.15 and the median increase equals 3. This suggests that the median increaser firm makes only 3 additional forward-looking profit statements in 1999 compared to 1996.

Table 7.3. Distribution of Disclosure Scores by Year

Panel A: Disclosure scores in 1996 (N=367)	
Average value	4.56
Minimum value	0
25th percentile	2
Median value	4
75th percentile	6
Maximum value	31
Standard deviation	3.94
Skewness	2.02
Panel B: Disclosure scores in 1999 (N=367)	
Average value	5.39
Minimum value	0
25th percentile	2
Median value	4
75th percentile	7
Maximum value	25
Standard deviation	4.30
Skewness	1.33

Table 7.3 presents the distribution of disclosure scores for the years 1996 and 1999. The definition of disclosure scores is based on the number of forward-looking profit sentences.

For the decreasers group, the firm with the most dramatic change in disclosure quality reduces the number of forward-looking profit sentences by 26. However, the median firm reduces the number of sentences only by 3. Overall, it is clear from Table 7.4 that changes in disclosure quality over time are, on average, relatively modest. This is consistent with the observation in the prior literature that disclosure levels over time are rather persistent.

Table 7.4. Distribution of Changes in Disclosure Scores: Profit Topics Only

Panel A: Disclosure increasers (N=188)	
Average value	4.15
Minimum value	1
25th percentile	2
Median value	3
75th percentile	5.5
Maximum value	18
Standard deviation	3.44
Skewness	1.73
Panel B: Disclosure decreasers (N=136)	
Average value	–3.50
Minimum value	–26
25th percentile	–5
Median value	–3
75th percentile	–1
Maximum value	–1
Standard deviation	3.26
Skewness	–3.49

Table 7.4 presents the distribution of changes in disclosure scores between 1996 and 1999. The definition of disclosure scores is based on the number of forward-looking profit statements.

7.5. Empirical Results

This section presents the empirical findings for the period 1996–1999. I start by reporting the results for the two groups of disclosure increasers and disclosure non–increasers. The latter group includes firms with decreasing disclosure scores and also a number of firms that maintain their disclosure scores between 1996 and 1999. Combining disclosure decreasers with firms that maintain their disclosure scores allows me to get comparable sample sizes (188 increasers vs. 179 non–increasers). For each group, the incremental contribution of future earnings to the full model R^2 in the first and the last year is estimated.

Table 7.5 presents the regression results in relation to R^2. Clearly, the empirical findings do not support the hypothesis of this chapter. According to this hypothesis, the change in the incremental R^2 for disclosure increasers should be higher than the change in the incremental R^2 for disclosure non–increasers. However, Table 7.5 indicates that

the ability of future earnings news to explain current returns decreases for the disclosure increasers and increases for the disclosure non-increasers. The incremental R^2 for firms that increase their disclosure levels between 1996 and 1999 decreases from 0.1290 in 1996 to 0.0779 in 1999. In contrast, for the disclosure non–increasers, the change in the incremental contribution of future earnings is positive, increasing from 0.0681 to 0.1961. These results are counter-intuitive. The increase in the number of forward-looking profit sentences results in less future earnings news being reflected in current return, while a reduction in forward-looking profit sentences allows more earnings news to be incorporated in prices.

Table 7.5. Incremental R^2: Increasers vs. Non–increasers (Sample Period: 1996–1999)

	1996	**1999**	**Δ Incremental R^2**
Panel A: Disclosure increasers (N=189)			
The simple return-earnings model R^2	0.0734	0.0777	
The full model R^2	0.2024	0.1556	
The incremental R^2 due to future earnings	0.1290	0.0779	-0.0511
Panel B: Disclosure non-increasers (N=178)			
The simple return-earnings model R^2	0.1290	0.0562	
The full model R^2	0.1971	0.2523	
The incremental R^2 due to future earnings	0.0681	0.1961	0.1280

The full model is:

$$R_t = b_0 + b_1 X_t + \sum_{k=1}^{2} b_{k+1} X_{t+k} + \sum_{k=1}^{2} b_{k+3} R_{t+k} + b_6 EP_{t-1} + b_7 AG_t$$

The simple return-earnings model is:

$$R_t = b_0 + b_1 X_t$$

The dependent variable is current period return, R_t. R_t, R_{t+1} and R_{t+2} are calculated as buy-and-hold returns (inclusive of dividends) over a 12-month period, starting four months after the end of the previous financial year. X_t, X_{t+1} and X_{t+2} are defined as earnings change deflated by price. Both current and future earnings changes are deflated by price at the start of the return window for period t. EP_{t-1} is defined as period t–1's earnings over price four months after the financial year-end of period t–1. AG_t is the growth rate of total book value of assets for period t. Disclosure scores are calculated based on forward-looking profit topics. Firms are sorted into two groups with increasing and non-increasing disclosure scores. The incremental R^2 measures the amount of future earnings information that is impounded in current returns.

In an attempt to assess the sensitivity of the results, firms that maintain their disclosure scores between 1996 and 1999 are deleted.[65] Therefore, firms are divided into disclosure increasers and disclosure decreasers.[66] Table 7.6 reports the results of the incremental R^2 for each group.

The results of Table 7.6 are similar to those reported in Table 7.5. The change in the incremental R^2 for the decreaser group is still positive, increasing from 0.0676 in 1996 to 0.2397 in 1999.

Table 7.6. Incremental R^2: Increasers vs. Decreasers (Sample Period: 1996–1999)

	1996	**1999**	**Δ Incremental R^2**
Panel A: Disclosure increasers (N=189)			
The simple return-earnings model R^2	0.0734	0.0777	
The full model R^2	0.2024	0.1556	
The incremental R^2 due to future earnings	0.1290	0.0779	-0.0511
Panel B: Disclosure Decreasers (N=135)			
The simple return-earnings model R^2	0.0970	0.0357	
The full model R^2	0.1646	0.2754	
The incremental R^2 due to future earnings	0.0676	0.2397	0.1721

The full model is:

$$R_t = b_0 + b_1 X_t + \sum_{k=1}^{2} b_{k+1} X_{t+k} + \sum_{k=1}^{2} b_{k+3} R_{t+k} + b_6 EP_{t-1} + b_7 AG_t$$

The simple return-earnings model is:

$$R_t = b_0 + b_1 X_t$$

The dependent variable is current period return, R_t. R_t, R_{t+1} and R_{t+2} are calculated as buy-and-hold returns (inclusive of dividends) over a 12-month period, starting four months after the end of the previous financial year. X_t, X_{t+1} and X_{t+2} are defined as earnings change deflated by price. Both current and future earnings changes are deflated by price at the start of the return window for period t. EP_{t-1} is defined as period t–1's earnings over price four months after the financial year-end of period t–1. AG_t is the growth rate of total book value of assets for period t. Disclosure scores are calculated based on forward-looking profit topics. Firms are sorted into two groups with increasing and decreasing disclosure scores. The incremental R^2 measures the amount of future earnings information that is impounded in current returns.

In summary, the results do not confirm the hypothesised relation between disclosure changes and prices leading earnings. The results are also not consistent with the findings of the time-series analysis in Lundholm and Myers (2002). Four possible

[65] In the cross-sectional analysis, firms with disclosure scores in the second and third quartiles are deleted. In the current chapter, however, the second and third quartiles are not deleted due to the small sample size.

[66] These classifications are similar to those reported in Lundhom and Myers (2002).

explanations could be given for these unexpected results. First, the time span between 1996 and 1999 may not be long enough to capture the effects of changes in the level of disclosure because disclosure quality is relatively stable over time (Miller and Piotroski, 2000). Second, year-specific effects might have affected the above results. Third, a classification based on the change in levels of disclosure might be too weak. Finally, the change in the return-earnings relation might be more sensitive to certain types of forward-looking information than others.

The next four sections discuss these explanations in more detail and provide suggestions for testing their validity. These tests include (1) extending the time span between the first year and the last year in the analysis, (2) choosing a different sample period, (3) redefining change in disclosure quality as a change between high, medium and low disclosure groups, and (4) re-categorising firms according to the change in the number of 'strong' sentences, where strong sentences are identified through labour-intensive reading.

7.6. Extending the Time Span

As noted earlier in this chapter, Lundholm and Myers (2002) measure the effect of changes in corporate disclosure over an average time span of seven years. They use the longest possible time span because previous research had observed that there are numerous short-term deviations from the general downward trend in the relation between current returns and current earnings. Furthermore, a firm's disclosure policy tends to be persistent and meaningful changes can take years. This limits the effective length of the time series and the power of the tests (Gelb and Zarowin, 2002).

Thus, the time span between the first and last year in the analysis needs to increase. To achieve this, the year 2000 is chosen instead of 1999 as the final year.[67] Therefore, the new sample is drawn from 1996 as the first year and 2000 as the last year.[68] The time span between the first and the last year is now four years compared with three years in the previous section.

Table 7.7 shows the regression results for this test. These results reveal that the incremental R^2s for each disclosure group are still not in line with the hypothesis in this chapter. While the change in the incremental R^2 for the group of disclosure increasers turns in the right direction when compared to the previous section, the change in the incremental R^2 is still much greater for the group of disclosure non-increasers than for the group of disclosure increasers. The changes in the incremental R^2 are –0.0075 and 0.1521, respectively. The hypothesis of this chapter predicts that the change is greater for the increasers than for the non-increasers.[69]

In summary, extending the sample period by one year does not appear to change the results very much. However, the time span in this section is still shorter than that in Lundholm and Myers (2002), four years versus seven years. Unfortunately, it is not possible to extend the original sample period by more than one year. If later years were selected as the last year, then a large number of observations would be lost due to the unavailability of future years' earnings and return variables.

[67] I am unable to extend the original sample by more than one year. This is because of the unavailability of a large number of future years' earnings and return variables at the time of the data collection.
[68] The total number of firms on *Dialog* in 2000 is 1598. This represents 508 financial firms and 1090 non-financial firms. Financial firms are excluded. The remaining firms are matched with non-financial firms in 1996. The matched firms are equal to 623 firms. Further observations are deleted due to changing year-ends (17 firms), missing variables (206 firms) and outliers (83 firms). The observations used in the analysis are 317 firms.
[69] When firms with no change in disclosure scores between 1996 and 2000 are removed from Panel B, then the findings lead to approximately similar results. The change in the incremental R^2 for the disclosure decreasers is now 0.0835.

Table 7.7. Incremental R^2: Increasers vs. Non-increasers: Sample Period: 1996–2000

	1996	2000	Δ Incremental R^2
Panel A: Disclosure increasers (N=208)			
The simple return-earnings model R^2	0.0787	0.0215	
The full model R^2	0.2374	0.1727	
The incremental R^2 due to future earnings	0.1587	0.1512	–0.0075
Panel B: Disclosure Non-increasers (N=109)			
The simple return-earnings model R^2	0.1341	0.1314	
The full model R^2	0.1821	0.3315	
The incremental R^2 due to future earnings	0.0480	0.2001	0.1521

The full model is:

$$R_t = b_0 + b_1 X_t + \sum_{k=1}^{2} b_{k+1} X_{t+k} + \sum_{k=1}^{2} b_{k+3} R_{t+k} + b_6 EP_{t-1} + b_7 AG_t$$

The simple return-earnings model is:

$$R_t = b_0 + b_1 X_t$$

The dependent variable is current period return, R_t. R_t, R_{t+1} and R_{t+2} are calculated as buy-and-hold returns (inclusive of dividends) over a 12-month period, starting four months after the end of the previous financial year. X_t, X_{t+1} and X_{t+2} are defined as earnings change deflated by price. Both current and future earnings changes are deflated by price at the start of the return window for period t. EP_{t-1} is defined as period t–1's earnings over price four months after the financial year-end of period t–1. AG_t is the growth rate of total book value of assets for period t. Disclosure scores are calculated based on forward-looking profit topics. Firms are sorted into two groups with increasing and non-increasing disclosure scores. The incremental R^2 measures the amount of future earnings information that is impounded in current returns.

7.7. Another Sample Period

The unexpected results in the previous two sections may also be due to some uncharacteristic features of the involved years, especially 1996, which is the first year in both time-series analyses. To test this argument, the results of a further time-series analysis are reported in this section. The regression results in this section are based on the years 1997 (first year) and 2000 (last year).[70]

The regression results are reported in Table 7.8. The results are no longer counter-intuitive in that the change in the incremental R^2 for the non-increasers is higher than the corresponding change for disclosure increasers, however, they are not supportive of the hypothesis of this chapter, either. The changes in the incremental R^2 are close to

[70] The matched sample for this sample period comprises 873 firms. Two firms are deleted due to the unavailability of *Datastream* accounting and return data. Further observations are deleted as a result of missing variables (317 firms) and outliers (110 firms). The sample used in the regression analysis is 444 firms.

zero for both groups. The R^2 increases marginally from 0.1361 to 0.1387 for the group of disclosure increasers and declines from 0.1129 to 0.1094 for the disclosure non-increasers. This set of results suggests that the importance of future earnings to current returns does not vary with the level of forward-looking profit statements.[71] The results in Table 7.8 also suggest that the empirical findings in relation to the change in the incremental R^2 vary substantially with the examined time period.

Table 7.8. Incremental R^2: Increasers vs. Non–increasers (Sample Period: 1997–2000)

Independent Variables	First Year 1997	Last Year 2000	Δ Incremental R^2
Panel A: Disclosure increasers (N=205)			
The simple return-earnings model R^2	0.0533	0.0186	
The full model R^2	0.1894	0.1573	
The incremental R^2 due to future earnings	0.1361	0.1387	0.0026
Panel B: Disclosure non-increasers (N=239)			
The simple return-earnings model R^2	0.0296	0.0899	
The full model R^2	0.1425	0.1993	
The incremental R^2 due to future earnings	0.1129	0.1094	–0.0035

The full model is:

$$R_t = b_0 + b_1 X_t + \sum_{k=1}^{2} b_{k+1} X_{t+k} + \sum_{k=1}^{2} b_{k+3} R_{t+k} + b_6 EP_{t-1} + b_7 AG_t$$

The simple return-earnings model is:

$$R_t = b_0 + b_1 X_t$$

The dependent variable is current period return, R_t. R_t, R_{t+1} and R_{t+2} are calculated as buy-and-hold returns (inclusive of dividends) over a 12-month period, starting four months after the end of the previous financial year. X_t, X_{t+1} and X_{t+2} are defined as earnings change deflated by price. Both current and future earnings changes are deflated by price at the start of the return window for period t. EP_{t-1} is defined as period t–1's earnings over price four months after the financial year-end of period t–1. AG_t is the growth rate of total book value of assets for period t. Disclosure scores are calculated based on forward-looking profit topics. Firms are sorted into two groups with increasing and non-increasing disclosure scores. The incremental R^2 measures the amount of future earnings information that is impounded in current returns.

One possible explanation for this lack of stability is that the scoring methodology developed in Chapters 4 and 5 does not do a very good job in the time-series analysis. It is possible that this methodology is not fine enough in a framework when small changes in disclosure scores determine a firm's classification as disclosure increaser or disclosure decreaser (see Table 7.4). This contrasts with the cross-sectional analysis

[71] If firms with unchanged disclosure scores are removed from Panel B, then the incremental R^2 for the disclosure decreasers increases from 0.1067 to 0.1342.

where high disclosure firms with a median score of nine are compared against firms with virtually zero disclosure (see Table 6.1).

To examine whether this argument holds, the classifications are modified in two ways. The results of these two classifications are reported in the following two sections.

7.8. An Alternative Classification for Disclosure Change

Based on the original sample period 1996–1999 this section tests the effect of corporate disclosure changes on prices leading earnings by using an alternative method for defining changes in disclosure quality. This method allocates firms according to their disclosure scores in 1996 and 1999 into three groups of low, medium and high disclosure firms. For each firm, I then compare the disclosure scores in 1996 and 1999. This comparison allows me to classify firms into three groups: disclosure increasers, disclosure decreasers and firms that maintain their disclosure status between 1996 and 1999. Firms are only defined as disclosure increasers if they move to a higher disclosure group between 1996 and 1999. This is the case, for example, when the firm's classification is changed from 'low' in 1996 to 'medium' or 'high' in 1999. Similarly, firms are only classified as disclosure decreasers if they move to a lower disclosure group between 1996 and 1999. This classification eliminates firms from the group of disclosure increasers and decreasers that only change their disclosure scores marginally. These firms are now defined as firms that maintain their levels of disclosure between 1996 and 1999. Such firms are excluded from the regression analysis that follows. In a certain way, the classification criterion in this section is stronger than that applied in earlier sections in this chapter. It is now comparable to the classification scheme in the cross-sectional analysis which eliminates firms in the second and third quartiles.

However, at the same time, this also reduces the number of observations used in the two regressions, a number that is not very large in the time-series analysis anyway.

Table 7.9 reports the number of firms that improve, maintain and reduce their status between 1996 and 1999. Firms above (below) the diagonal improve (reduce) their status over the three-year period.

Table 7.9. Alternative Classification for Disclosure Change

	Low 1999	Medium 1999	High 1999
Low 1996	76 = Maintain	58 = Increase	32 = Increase
Medium 1996	40 = Decrease	52 = Maintain	70 = Increase
High 1996	21 = Decrease	50 = Decrease	64 = Maintain

The regression results for the new classification scheme are reported in Table 7.10. As usual, Panel A reports the incremental R^2 for the disclosure increasers and Panel B indicates the findings for the group of disclosure decreasers. The regression results in Table 7.10 show that the change in the incremental R^2 is positive for the disclosure increasers as predicted. However, the change in the incremental R^2 is greater for the group of decreasers than for the group of increasers (0.0953 against 0.0771). This contrasts with the prediction in this chapter that the incremental R^2 should be greater for disclosure increasers than for disclosure decreasers. Overall, it appears that the modification of the classification scheme in this section has not changed the empirical findings very much.

Having been unable to find the predicted relation between prices leading earnings and automated disclosure scores in a time-series framework, the final test in this chapter abandons the automated scoring methodology altogether in favour of labour-intensive reading. In particular, the analysis in the following section examines whether the failure to find the predicted relation in a time-series framework is due to the inability of the

automated scoring system to differentiate between 'strong' and 'weak' trading sentences.

Table 7.10. Incremental R^2: Changes in Disclosure Quality Based on Disclosure Ranking (Sample Period: 1996–1999)

	1996	1999	Δ Incremental R^2
Panel A: Disclosure increasers (N=151)			
The simple return-earnings model R^2	0.0902	0.0531	
The full model R^2	0.0844	0.1244	
The incremental R^2 due to future earnings	-0.0058	0.0713	0.0771
Panel B: Disclosure decreasers (N=101)			
The simple return-earnings model R^2	0.0507	0.0145	
The full model R^2	0.1655	0.2246	
The incremental R^2 due to future earnings	0.1148	0.2101	0.0953

The full model is:

$$R_t = b_0 + b_1 X_t + \sum_{k=1}^{2} b_{k+1} X_{t+k} + \sum_{k=1}^{2} b_{k+3} R_{t+k} + b_6 EP_{t-1} + b_7 AG_t$$

The simple return-earnings model is:

$$R_t = b_0 + b_1 X_t$$

The dependent variable is current period return, R_t. R_t, R_{t+1} and R_{t+2} are calculated as buy-and-hold returns (inclusive of dividends) over a 12-month period, starting four months after the end of the previous financial year. X_t, X_{t+1} and X_{t+2} are defined as earnings change deflated by price. Both current and future earnings changes are deflated by price at the start of the return window for period t. EP_{t-1} is defined as period t–1's earnings over price four months after the financial year-end of period t–1. AG_t is the growth rate of total book value of assets for period t. Disclosure scores are calculated based on forward-looking profit topics. Firms are classified as disclosure increasers (decreasers) if the disclosure classification improved (reduced) between 1996 and 1999. The incremental R^2 measures the amount of future earnings information that is impounded in current returns.

7.9. Identifying Strong Sentences through Labour-Intensive Reading Definitions

A major contribution of the thesis is the development of a scoring methodology that partially automates the generation of disclosure scores and thus allows the researcher to produce disclosure scores for large samples of firms. While a scoring system based on text searches is cost-efficient and easy to replicate, it is unlikely to be as accurate as labour-intensive reading. This section examines whether my inability to find the predicted relation between changes in disclosure scores and the importance of future earnings for current prices is due to a degree of noise that is inevitably introduced into disclosure scores when scores are generated via a text search approach. To investigate this issue all sentences identified by *Nudist* as forward-looking profit statements are

carefully read. I then decide whether these sentences are 'strong' trading statements about the future or not.[72]

A sentence is classified as a 'strong' statement (1) if the information is related directly to margins, operating profits or bottom-line profits and (2) if the time horizon is clearly specified. The following three sentences are examples of forward-looking profit statements that I judge to be 'strong' trading statements about the future:

> **Sentence (1):** *We **anticipate profits** in the first year of between $600,000 and $1million.*' (Druck Holdings PLC – 1996 Annual Report and Accounts)
>
> **Sentence (2):** *'Inflation, which has persisted at high levels in recent years, is trending downwards and this, allied to increased output resulting from adequate water supplies, should result in a significant improvement in operating **profits** in **1997**.'* (Lonrho PLC – 1996 Annual Report and Accounts)
>
> **Sentence (3):** *'We are anticipating making an operating **profit** in France in **1997** and will be considering plans to develop the business profitably towards national coverage.'* (Brake Bros PLC – 1996 Annual Report and Accounts)

All three sentences make statements about next year's (operating) profit. I exclude sentences that refer to the past year's profit. The following sentences are examples of forward-looking profit statements that I consider as not strong trading statements about the future:

Examples: The firm's past year profit

> **Sentence (1):** *'This **profit forecast** was achieved.'* (IMI Plc– 1999 Annual Report and Accounts)
>
> **Sentence (2):** *'On **sales** of £12.5 million (1998: £15.1 million) the pre-tax **loss** was in line with **forecast** at £639,000 (1999: pre-tax profit £454,000).'* (Radamec Plc– 1999 Annual Report and Accounts)

In addition, I exclude sentences that do not directly relate to profits, but rather to the firm's environment and/or the firm's industry. The following sentences are examples of forward-looking profit statements that I consider as not strong trading statements about the future:

[72] The alternative to checking all forward-looking profit statements identified by *Nudist* would be to read the entire annual report discussion section from scratch. Unfortunately, with 2 x 367 = 734 firm-years, this was regarded as too time-consuming.

Example: The firm's environment

> *'The **trading** environment in which the Construction Divisions operate **remains** harsh.'* (MJ Gleeson Group Plc – 1996 Annual Report and Accounts)

Example: The firm's industry

> *'The outdoor advertising industry has seen notable consolidation in the UK and a smaller number of larger players **is expected** to bring further **benefits** to the profile and organisation of the outdoor industry.'* (Scottish Radio Holdings Plc – 1999 Annual Report and Accounts)

While economy-wide factors and trends in the industry and product market obviously have implications for future profit, the relation between these trends and profits are much vaguer.

Furthermore, I exclude from the scores any forward-looking profit statements in which the time horizon is not well defined. The following statement is an example of a forward-looking profit sentence that I consider as not a strong trading statement:

> *'This strategy **will**, we believe, generate high **returns** for our shareholders over the long term.*' (Electrocomponents Plc – 1999 Annual Report and Accounts)

Moreover, I exclude forward-looking profit statements when their associated accounting topics have different meanings. The example below clarified this idea:

> *'It too has been given a new livery and **will return** to television advertising in the new financial year.'* (HP Bulmer Holdings Plc – 1999 Annual Report and Accounts)

The above example contains a profit topic (the keyword return) and a well-specified time horizon (the keyword: the new financial year); however, the keyword 'return' in this statement has a different meaning. It does not refer to the accounting return; it means in this context 'going back'.

Finally, for the purpose of calculating the scores, I exclude sentences that contain thankful words to directors, employees, suppliers or customers and sentences that – despite including a forward-looking key work and a profit-related topic – do not appear

to have any profit implications whatsoever. The following sentence is an example of a 'forward-looking profit statement' that is not considered as a strong trading statement:

> *'I should like to thank Sir Terence Higgins, who **will** be retiring at the Annual General Meeting in March 1997, for his invaluable **contribution** to the Group over the past six years'.* (First Choice Holidays PLC - 1996 Annual Report an Accounts)

The analysis in this section is once again based on the years 1996 and 1999, with the change in disclosure scores being determined for 367 firms. The total number of sentences read is 1674 for 1996 and 1978 for 1999. As usual, separate regressions are run for the group of disclosure increasers and disclosure non-increasers. The results are reported in Panels A and B of Table 7.11.

The new classification scheme, which is now based on a mixture of automation and manual reading, leads to empirical findings that are consistent with the hypothesis of this chapter. The change in incremental R^2 is higher for the disclosure increasers than for the disclosure non-increasers (0.0840 against –0.1110). While the amount of future earnings news that is reflected in current returns increases for the group of disclosure increasers, it declines for the group of firms that reduces (or maintains) the number of strong forward-looking profit statements. This finding is consistent with forward-looking profit statements providing credible information to the stock market about future trading performance.[73]

[73] When firms with the same level of strong forward-looking profit statements are excluded from the analysis, then the change in the incremental R^2 for the group of disclosure decreasers is –0.2398, producing further evidence for the success of the classification scheme in this section.

Table 7.11. Incremental R^2: 'Strong' Forward-looking Profit Sentences (Sample Period: 1996–1999)

	1996	1999	Δ Incremental R^2
Panel A: Disclosure increasers (N=246)			
The simple return-earnings model R^2	0.0923	0.1122	
The full model R^2	0.1909	0.2948	
The incremental R^2 due to future earnings	0.0986	0.1826	0.0840
Panel B: Disclosure non-increasers (N=121)			
The simple return-earnings model R^2	0.1241	0.0235	
The full model R^2	0.2869	0.0753	
The incremental R^2 due to future earnings	0.1628	0.0518	–0.1110

The full model is:

$$R_t = b_0 + b_1 X_t + \sum_{k=1}^{2} b_{k+1} X_{t+k} + \sum_{k=1}^{2} b_{k+3} R_{t+k} + b_6 EP_{t-1} + b_7 AG_t$$

The simple return-earnings model is:

$$R_t = b_0 + b_1 X_t$$

The dependent variable is current period return, R_t. R_t, R_{t+1} and R_{t+2} are calculated as buy-and-hold returns (inclusive of dividends) over a 12-month period, starting four months after the end of the previous financial year. X_t, X_{t+1} and X_{t+2} are defined as earnings change deflated by price. Both current and future earnings changes are deflated by price at the start of the return window for period t. EP_{t-1} is defined as period t–1's earnings over price four months after the financial year-end of period t–1. AG_t is the growth rate of total book value of assets for period t. Disclosure scores are calculated based on forward-looking profit topics. Firms are sorted into two groups with increasing and non-increasing their number of strong forward-looking profit sentences. The incremental R^2 measures the amount of future earnings information that is impounded in the current return.

The results of the time-series analysis in this chapter are important because they suggest that, while the automated text search approach developed in this thesis appears to work well in a cross-sectional setting, it seems unable to pick up subtle differences in disclosure quality that appear important in a time-series framework where the amount of disclosure changes relatively little. With hindsight, it appears that an important feature of the cross-sectional analysis is that it effectively compares firms at the top end of the disclosure spectrum against firms with virtually no disclosure at all. If more subtle differences in the level of disclosure matter – like in a time-series analysis with relatively persistent disclosure scores over time – then the automated text search approach appears to be less effective. This is an important qualification of the scoring system developed in this thesis.

7.10. Summary

To complement the cross-sectional results in Chapter 6, the current chapter follows Lundholm and Myers's (2002) methodology to examine the association between changes in voluntary disclosures and changes in the extent to which returns anticipate future earnings changes.

The hypothesis in the present chapter predicts that the change in the incremental R^2 for disclosure increasers is greater than the change in the incremental R^2 for disclosure decreasers. However, initially the empirical findings are not in line with the hypothesised prediction.

In an attempt to examine possible explanations, several modifications of the original research design are considered. First, the time span is extended between the first and the last year in the analysis. Second, a new sample period is chosen. Third, disclosure change is redefined according to the change in the disclosure rank for each firm in the first and the last year of the analysis. The results of these modifications are generally not consistent with the prediction that the incremental R^2 contributed by future earnings news increases for disclosure increasers relative to disclosure decreasers. Finally, I re-categorise firms according to the change in the number of 'strong' trading statements identified through manual reading. The results of this final modification provide evidence that changes in forward-looking disclosure is positively related to changes in the amount of future earnings news reflected in current returns.

Chapter 8: Conclusion

The present study aims to test the ability of partially automated disclosure scores for identifying the information content of annual report narratives for future earnings. This chapter provides a summary of the main findings followed by the main implications of these findings. It also discusses the limitations of the study. The chapter ends by suggesting several avenues for future research.

8.1. Summary

This study investigates the impact of forward-looking disclosure levels in annual report narratives on prices leading earnings. The findings are based on large samples of annual reports electronically available in the *Dialog* database. 800 annual reports per year for the time period 1996-1999 were collected. *Nudist* was used to score annual report narratives, and then I examined the association between disclosure quality and prices leading earnings using the Collins et al. (1994) regression model.

To measure the quality of corporate disclosure, I used a new methodology for constructing the list of disclosure items. Novel features of this methodology were related to the use of analysts' reports to construct the list of disclosure items and the use of *Nudist* to speed up the process of constructing this list. The scoring sheet is based on the forward-looking disclosure topics that are included in a sample of 60 analysts' reports. The scoring sheet comprises a list of 500 topics that are important for financial analysts in their forecasting process.

To score large samples of annual reports, a new methodology was developed. The scoring procedure generally followed the following stages. First, I identified a list of disclosure topics from analyst reports. Second, I identified a list of forward-looking key words that are frequently used in the annual report. Third, *Nudist* was used to run two

separate text-searches. The first search identified the number of forward-looking sentences in each annual report. The second search identified the number of sentences that included at least one specific topic. Finally, *Nudist* formed the intersection between the two searches. The overall results of the scoring methodology were exported to a spreadsheet, which included the total number of forward-looking sentences with a relevant topic for each firm.

The reliability of the disclosure scores was supported by providing a clear statement of procedures and a clear definition of forward-looking information. Therefore, other researchers should be able to replicate the results of this study easily. All that is needed is a list of forward-looking key words and a topic list.

The validity of the disclosure scores was supported by three different sets of analyses: (1) The correlation between disclosure and firm characteristics identified in prior research to be associated with disclosure level, (2) the correlation among the components of the scoring sheet, and (3) the correlation between the scores produced by the techniques deployed in this study and the scores produced by a manual based content analysis methodology.

After calculating the disclosure scores for large samples of firms, the main aim then was to test the association between forward-looking disclosure in annual report narratives and prices leading earnings. The cross-sectional regression results in Chapter 6 were unable to find a significant relation between forward-looking disclosure based on an all topic list and prices leading earnings. These results help to explain why Gelb and Zarowin (2002) do not find a positive relation between annual report disclosures and

'prices leading earnings'.[74] When disclosure scores were defined as the number of forward-looking profit sentences, the regression results show that the quality of disclosure is significantly associated with prices leading earnings. This effect is, however, only significant for the one-year-ahead earnings changes.

These results support the view that current earnings alone have only a limited ability to communicate a firm's value to the market. The market uses additional disclosures to anticipate future earnings.

These results also show that the use of the automated disclosure scores is useful for measuring the association between annual report narratives and prices leading earnings. This is true when firms at the top end of the disclosure spectrum are compared against firms with virtually no disclosure at all.

The second set of results considered the association between changes in voluntary corporate disclosures and changes in the extent to which returns anticipate future earnings changes. I followed the methodology in Lundholm and Myers (2002). The empirical results for the sample period 1996-1999 are not in line with the hypothesised prediction. Four modifications were made to the original research design. These included extending the time-span, selecting another sample period, redefining disclosure changes according to their ranks in the first and the last year, and re-categorising firms according to the change in the number of 'strong' sentences.

The findings of the first three modifications were generally not consistent with the hypothesised prediction. Therefore, the decision was made to use a labour-intensive approach beside the computerised scoring method to identify changes of disclosure

[74] AIMR ratings for the annual report are based on quantitative and qualitative information, both forward-looking and backward looking. Thus, the resulting scores are likely to contain too much noise for identifying prices leading earnings.

scores for each firm. All sentences identified by *Nudist* as forward-looking profit statements were carefully read. I then identified 'strong' sentences that explicitly referred to profit topics for a specific period of time. The results of this modification provide evidence that changes in a firm's forward-looking disclosure is positively related to changes in the amount of future earnings news reflected in current returns. Thus it appears that the effect is strongest if the information is related directly to margins, operating profits or bottom-line profits and if the time horizon is clearly specified.

The results of the time-series analysis are important because they suggest that – while the automated text search approach developed in this thesis appears to work well in a cross-sectional analysis – it seems not to able to pick up subtle differences in disclosure quality that appear important in a time-series framework where the amount of information changes are relatively small. If more subtle differences in the level of disclosure matter – like in a time-series analysis with relatively persistent disclosure scores over time – then the automated text search approach appears to be less effective. This is an important qualification of the scoring system developed in this thesis.

In summary, the present study showed that using *Nudist* as a more objective measurement of arriving at disclosure scores is a powerful tool for analysing very large quantities of textual data as covered in Chapter 6. This is consistent with Holsti (1969) who suggests that researchers should use computers in qualitative research when their analysis is extremely complex and is based on large numbers of text. On the other hand, when the sample sizes are relatively small, the use of a less-objective measurement is more appropriate – as shown in Chapter 7 where the association between changes in disclosures and changes in the return-earnings association was found to be significant.

These results suggest that combining computer methodology with manual reading methodology may provide the basis for larger sample studies.

8.2. Implications

There are many reasons for undertaking this study. The most important is the fact that this type of research has potential implications. It helps to inform regulators about the benefits of corporate voluntary disclosures to investors and the disclosing firm. The study provides evidence that forward-looking profit information, published in annual report narratives, is useful to investors in predicting future earnings changes. As a result, this study provides part of the information that is needed for a more informed cost-benefit analysis of increased disclosures.

The findings reported in this study also have managerial implications. They show that markets are uncertain about the quality of reported earnings and that the market appears to be better informed when including forward-looking profit information in annual report narratives. Therefore, for effective financial communication with the stock market, managers should give high priority to develop appropriate and complete disclosure practices. The findings reported in the study provide assistance to managers wishing to understand more precisely how forward-looking disclosures affect earnings' timeliness. In particular, the results demonstrate that including profit forecasts in annual report narratives enables investors to anticipate future earnings more accurately.

In addition, the findings of the study have important implications for small investors who may not have access to information through other sources in the same way that financial analysts or large institutional investors do. The results reported in the present study suggest that reported earnings alone may be insufficient for an investor to anticipate future earnings changes. In addition, they suggest that forward-looking profit

information improves investors' earnings forecasts and then guides their investment decisions.

Finally, the present study has implications for the efficient market hypothesis. The results of the study suggest that increased levels of forward-looking profit disclosures in annual report narratives provide investors with value-relevant information. This information enables them to better anticipate future earnings. Accordingly, this leads to more efficient capital markets.

8.3. Limitations

The use of the computerised content analysis approach has enabled me to produce disclosure scores for very large samples of annual reports. Hence, it enabled me to undertake a large scale disclosure study. On the other hand, the present study suffers from a number of limitations. First, the study assumes that the set of value relevant disclosure topics does not vary across industrial sectors. However, it appears that this assumption may not be proper. In particular, prior studies highlight the importance of specific disclosure topics to specific industries. For example, Nagar and Rajan (2001) find that disclosures related to 'defects' and 'on-time delivery' measures are thought to be value-relevant in the hotel sector.

Second, the study calculates disclosure scores by simply adding up the number of sentences. This approach ignores the fact that the usefulness of disclosures can vary from sentence to sentence, and this limitation was particularly obvious in the time-series analysis.

Finally, the study ignores the importance of corporate governance mechanisms such as board structure and ownership structure. It also ignores the effect of corporate

proprietary cost. Such characteristics are potentially important inputs when assessing the adequacy of corporate disclosures.

8.4. Suggestions for Future Research

The present study suggests a number of other avenues for future research. First, a random sample of 60 analysts' reports was selected to construct the list of disclosure topic. These reports covered companies from different sectors. This study does not consider the importance of some disclosure topics to specific sectors. The most obvious example is the importance of the research and development (R&D) to the pharmaceutical industry. Further research could be carried out to construct a more specific scoring sheet. This sheet would show disclosure items by each industry sector. In this case, a higher disclosure score could be given to a disclosure item that is more important for a specific industry.

Second, in prior research, quantitative management forecasts have been used as a proxy for corporate disclosure. These studies find that quantitative forecasts have value-relevant information. In particular, they find that quantitative management forecasts are positively associated with the accuracy of analyst forecasts. It would be interesting to investigate the importance of such specific forecasts in the UK. In particular, one can replicate the scoring methodology adopted in this thesis and attach a greater weight to quantitative management forecasts than to other items.

Third, I believe that the process of analysing annual report narratives has scope for further refinement. Currently, my methodology equates disclosure quality with the amount of information provided. In contrast, identifying the underlying tone of voluntary disclosures means being able to discriminate between good news and bad

news and such a refinement could be extremely useful for studying the benefits of corporate voluntary disclosures.

Fourth, the validity of my disclosure scores is measured in several ways. One of these measures is the correlation between the scores produced by *Nudist* and those produced by a labour-intensive approach. Future research could also make a direct comparison of the scores produced by *Nudist* and the AIMR-FAF ratings. Whilst the latter are only available for a sample of US firms, it might nevertheless be of interest to study the properties of the two types of disclosure rankings side by side.

Fifth, the study highlights the importance of automated disclosure scores for prices leading earnings. Additional research could be undertaken to examine other contexts such as cost of equity capital, cost of debt capital, analyst following and the characteristics of analysts' forecasts.

Sixth, in Chapter 7 it was noticed that extending the sample period further by one year does not appear to improve the association between disclosure changes and the relation between return and future earnings. This is because a firm's disclosure policy tends to be persistent and meaningful changes can take years. This limits the effective length of the time series and the power of the tests. It was not possible to extend the original sample by more than one year due to the unavailability of future years' earnings and return variables at the time of the data analysis. It would be interesting to re-examine the association between changes in the UK corporate disclosure practice and changes in earnings' timeliness over a longer time span.

Seventh, a labour-intensive approach was used to identify value-relevant disclosures in Section 7.9. Using a semi-computerised content analysis approach provides me significant results. It is possible that the involvement of experts in linguistics in

determining further key words (in addition to the forward-looking key words and the topic list) may improve the ability of my computer-based scoring methodology for identifying value-relevant disclosures. However, the potential contribution from the application of additional linguistic methods remains an area for future research as it possible that there will be significant difficulties in overcoming some of the classificatory problems of, for example, forward/backward looking sentences.

Eight, previous studies show that earnings are likely to be relatively less informative when a firm reports losses. Because losses cannot continue indefinitely, they are a poor predictor of future earnings (Hayn, 1995). It would be interesting to study the association between corporate disclosures and prices leading earnings for loss-making firms that (1) do not change their disclosure level and (2) increase (decrease) their disclosure level over a specific period of time.

Finally, a large number of studies provide evidence on the association between the levels of corporate voluntary disclosure and a set of corporate governance mechanisms such as ownership structure and board composition (e.g., Eng and Mak, 2003). Therefore, it would be interesting to extend the present study by testing the extent to which these mechanisms affect the association between disclosure quality and the informativeness of stock prices for future earnings.

References

Abdolmohammadi, M., Simnett, R., Thibodeau, J. and Wright, A. (2003). 'Sell-side financial analysts' quarterly reports and opportunities for assurance services'. Working Paper, Bentley College, USA.

Abrahamson, E. and Amir, E. (1996). 'The information content of the president's letter to shareholders'. *Journal of Business Finance & Accounting*, 23 (8): 1157–1182.

Ahmed, K. and Courtis, J. K. (1999). 'Associations between corporate characteristics and disclosure levels in annual reports: a meta-analysis'. *The British Accounting Review*, 31 (1): 35–61.

Amihud, Y. and Mendelson, H. (1986). 'Asset pricing and the bid-ask spread'. *Journal of Financial Economics*, 17 (2): 223–249.

Anandarajan, A., Kleinman, G. and Palmon, D. (2000). 'Investors' expectations and the corporate information disclosure gap: a perspective'. *Research in Accounting Regulation*, 14: 245–260.

Arnold, J. and Moizer, P. (1984). 'A survey of the methods used by UK investment analysts to appraise investments in ordinary shares'. *Accounting and Business Research*, 14 (55): 195–207.

Baginski, S. and Hassell, J. (1990). 'The market interpretation of management earnings forecasts as a predictor of subsequent financial analyst revision'. *The Accounting Review*, 65 (1): 175–190.

Ball, R. and Brown, P. (1968). 'An empirical evaluation of accounting income numbers'. *Journal of Accounting Research*, 6 (2): 159–178.

Barker, R. G. (1998). 'The market for information – evidence from finance directors, analysts and fund managers'. *Accounting and Business Research*, 29 (1): 3–20.

Barron, O. E., Kile, C.O. and O'Keefe, T. B. (1999). 'MD&A quality as measured by the SEC and analysts' earnings forecasts'. *Contemporary Accounting Research*, 16 (1): 75–109.

Barron, O. E., Kile, C. and Venkataraman, P. (2001). 'The incremental price effects of SEC mandated MD&A disclosure'. Working Paper, Pennsylvania State University.

Basu, S. (1997). 'The conservatism principle and the asymmetric timeliness of earnings'. *Journal of Accounting and Economics*, 24 (1): 3 –37.

Beattie, V., McInnes, B. and Fearnley, S. (2001). 'The analysis of narratives in annual reports: a multidimensional framework'. Financial Reporting and Business Communication Conference, Cardiff Business School, 5th-6th July.

Beattie, V., McInnes, B. and Fearnley, S. (2002). *Through the Eyes of Management: A Study of Narrative Disclosures.* Research Report, London: ICAEW.

Beattie, V. and Pratt, K. (2002). *Voluntary Annual Report Disclosures: What Users Want*, Research Report. Edinburgh: ICAS.

Beaver, W. H. (1968). 'The information content of annual earnings announcements'. *Journal of Accounting Research*, 6 (Supplement): 67–92.

Beaver, W. H., Lambert, R. and Morse, D. (1980). 'The Information content of security prices'. *Journal of Accounting and Economics*, 2 (1): 3–28.

Beaver, W. H., Lambert, R. and Ryan, S. (1987). 'The information content of security prices: a second look'. *Journal of Accounting and Economics*, 9 (2): 139–157.

Berelson, B. (1952). *Content Analysis in Communication Research*. Glencoe, Ill: Free Press.

Biggs, S. F. (1984). 'Financial analysts' information search in the assessment of corporate earning power'. *Accounting, Organisations and Society*, 9 (3/4): 313–323.

Blasco, M. E. and Trombetta, M. (2003). 'The reputational consequences of better disclosures quality'. Working Paper, Third Accounting Research Workshop SIMT, Stuttgart, Universidad Carlos III de Madrid.

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

Botosan, C. A. and Harris, M. S. (2000). 'Motivations for a change in disclosure frequency and its consequences: An examination of voluntary quarterly segment disclosures'. *Journal of Accounting Research*, 38 (2): 329–353.

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

Bouwman, M. J., Frishkoff, P. A. and Frishkoff, P. (1987). 'How do financial analysts make decisions? a process model of the investment screening decision'. *Accounting, Organisations and Society*, 12 (1): 1–30.

Bowen, R., Davis, A. and Matsumoto, D. (2002). 'Do conference calls affect analysts' forecasts?' *The Accounting Review*, 77 (2): 285–316.

Bradshaw, M. T. (2002). 'The use of target prices to justify sell-side analysts' stock recommendations'. *Accounting Horizons*, 16 (1): 27–41.

Breton, G. and Taffler, J. (2001). 'Accounting information and analyst stock recommendation decisions: a content analysis approach'. *Accounting and Business Research*, 31 (2): 91–101.

Bricker, R., Previts, G., Robinson, T. and Stephen, Y. (1995). 'Financial analyst assessment of company earnings quality'. *Journal of Accounting, Auditing and Finance*, 10 (3): 541–554.

Brown, L. and Han, J. (1992). 'The impact of annual earnings announcements on convergence of beliefs'. *The Accounting Review*, 67 (4): 862–875.

Brown, P. R. (1997). 'Financial data and decision making by sell-side analysts'. *The Journal of Financial Statement Analysis*, Spring: 43–48.

Brown, P. (2001). *Capital Markets-Based Research in Accounting: An Introduction*. Monograph, Coopers and Lybrand (Australia) and Accounting Association of Australia and New Zealand, Melbourne.

Brown, S., Finn, M. and Hillegeist, S. (2001). 'Disclosure quality and the probability of informed trade'. Working Paper, Northwestern University.

Brown, P., Taylor, S. and Walter, T. (1999). 'The impact of statutory sanctions on the level and information content of voluntary corporate disclosure'. *Abacus*, 35 (2): 138–162.

Bryan, S. H. (1997). 'Incremental information content of required disclosures contained in management discussion and analysis'. *The Accounting Review*, 72 (2): 285–301.

Bujaki, M., Zeghal, D. and Bozec, R. (1999). 'The disclosure of future oriented information in annual reports of Canadian corporations'. Working Paper: 44, University of Ottawa, Canada.

Bushee, B. and Noe, C. (2000). 'Corporate disclosure practices, institutional investors and stock return volatility'. *Journal of Accounting Research*, 38 (Supplement): 171–202.

Butler, M., Kraft, A. and Weiss, I. S. (2003). 'The effect of reporting frequency on the timeliness of earnings: the cases of voluntary and mandatory interim reports'. Working Paper: FR02–11, Simon Business School, University of Rochester.

Byard, D. and Shaw, K. (2003). 'Corporate disclosure quality and properties of analysts' information environment'. *Journal of Accounting, Auditing and Finance*, 18 (3): 355–378.

Carner, T. (1972). *Content Analysis: A Technique for Systematic Inference for Communications*. London, UK, B T Batsford.

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

Chandra, G. (1975). 'Information needs of security analyst'. *The Journal of Accountancy*, December: 65–70.

Chang, L. S. and Most, K. S. (1985). *The Perceived Usefulness of Financial Sentences for Investors' Decisions. University Presses of Florida*, Florida International University Press, Miami, FL.

Cho, J. and Jung, K. (1991). 'Earnings response coefficients: a synthesis of theory and empirical evidence'. *Journal of Accounting Literature,* 10: 85–116.

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

Clarkson, P. M., Dontoh, A., Richardson, G. D. and Sefcik, S. (1992). 'The voluntary inclusion of earnings forecasts in IPO prospectuses'. *Contemporary Accounting Research,* 8 (2): 601–626.

Clarkson, P. M., Kao, J. L. and Richardson, G. D. (1994). 'The voluntary inclusion of forecasts in the MD&A section of annual reports'. *Contemporary Accounting Research,* 11 (1–II): 423– 450.

Clarkson, P., Guedes, J. and Thompson, R. (1996). 'On the diversification, observability and measurement of estimation risk'. *Journal of Financial and Quantitative Analysis,* 31 (1): 69–84.

Clarkson, P. M., Kao, J. L. and Richardson, G. D. (1999). 'Evidence that management discussion and analysis (MD&A) is a part of a firm's overall disclosure package'. *Contemporary Accounting Research,* 16(1): 111–134.

Clatworthy, M. and Jones, M. J. (2003). 'Financial reporting of good news and bad news: evidence from accounting narratives'. *Accounting and Business Research,* 33 (3): 171–185.

Coles, J. L., Loewenstein, U. and Suay, J. (1995). 'On equilibrium pricing under parameter uncertainty'. *Journal of Financial and Quantitative Analysis,* 30 (3): 347–364.

Collins, D. W. and Kothari, S. P. (1989). 'An analysis of intertemporal and cross-sectional determinants of earnings response coefficients'. *Journal of Accounting and Economics,* 11 (2–3): 143–181.

Collins, D. W., Kothari, S. P., Shanken, J. and Sloan, R. G. (1994). 'Lack of timeliness and noise as explanations for the low contemporaneous return-earnings association'. *Journal of Accounting and Economics,* 18 (3): 289–324.

Cooke, T. E. (1989a). 'Disclosure in the corporate annual reports of Swedish companies'. *Accounting and Business Research,* 19 (74), 113–124.

Cooke, T. E. (1989b). 'Voluntary corporate disclosure by Swedish companies'. *Journal of International Financial Management and Accounting,* 1 (2): 171–195.

Cooke, T. E. (1991). 'An assessment of voluntary disclosure in the annual reports of Japanese corporations'. *The International Journal of Accounting,* 26 (3): 174–189.

Core, J. E. (2001). 'A review of the empirical disclosure literature: discussion'. *Journal of Accounting and Economics,* 31 (3): 441–56.

Day, J. F. (1986). 'The use of annual reports by UK investment analysts'. *Accounting and Business Research*, 16 (64): 295–307.

Diamond, D. W. and Verrecchia, R. E. (1991). 'Disclosure, Liquidity, and the Cost of Capital'. *Journal of Finance*, 46 (4): 1325–1359.

Donnelly, R. and Walker, M. (1995). 'Share price anticipation of earnings and the effect of earnings persistence and firm size', *Journal of Business Finance and Accounting*, 22 (1): 5–18.

Doula, D. (2000). *Management's Forecasts and Share Price Anticipation of Earnings: UK Evidence from the Building and Construction, Building Materials and Merchants, Leisure and Hotels and Media Industry Sectors*. MSc Dissertation, UMIST.

Dye, R. A. and Sridhar, S. S. (2002). 'Resource allocation effects of price reactions to disclosures'. *Contemporary Accounting Research*, 19 (3): 385–410.

Eng, L. L. and Mak, Y. T. (2003). 'Corporate governance and voluntary disclosure'. *Journal of Accounting and Public Policy*, 22(4): 293–376.

Eng, L. L. and Teo, H. K. (2000). 'The relation between annual report disclosures, analysts' earnings forecasts and analyst following: Evidence from Singapore'. *Pacific Accounting Review*, 11 (2): 219–239.

Epstein, M. J. and Palepu, K. (1999). 'What financial analysts want?' IR Update – National Investor Relations Institute Monthly Newsletter, May, 6–10.

Field, L., Lowry, M. and Shu, S. (2003). 'Does disclosure deter or trigger litigation?'. Working Paper, MSAF Research Seminar, Manchester University.

Firth, M. (1980). 'Raising finance and firm's corporate reporting policies'. *Abacus*, 16 (2): 100–115.

Foster, G. J. (1986). *Financial Statement Analysis*, 2nd edition, Prentice Hall.

Francis, J., Hanna, J. D. and Philbrick, D. R. (1997). 'Management communications with securities analysts'. *Journal of Accounting and Economics*, 24 (3): 363–394.

Frankel, R., Johnson, M. F. and Skinner, D. (1999). 'An empirical investigation of conference calls as a voluntary disclosure medium'. *Journal of Accounting Research*, 37 (1): 133–150.

Frazier, K. B., Ingram, R.W. and Tennyson, B.M. (1984). 'A methodology for the analysis of narrative accounting disclosures'. *Journal of Accounting Research*, 22 (1): 318–331.

Freeman, R. (1987). 'The Association between accounting earnings and security returns for large and small firms'. *Journal of Accounting and Economics*, 9 (2): 195–228.

Frost, C. A. (2002). 'Characteristics and information value of corporate disclosures of forward-looking information in global equity markets', Working Paper, Sawyer School of Management, Suffolk University.

Gelb, D. S. and Zarowin, P. (2002). 'Corporate disclosure policy and the informativeness of stock prices'. *Review of Accounting Studies*, 7 (1): 33–52.

Glosten, L., and Milgrom, P. (1985). 'Bid, ask, and transaction prices in a specialist market with heterogeneously informed traders'. *Journal of Financial Economics*, 26: 71–100

Govindarajan, V. (1980). 'The objective of financial sentences: an empirical study of the use of cash flow and earnings by security analysts'. *Accounting, Organisations and Society*, 5 (4): 383–392.

Gray, S. J., Meek, G. K. and Roberts, C. B. (1995). 'International capital market pressures and voluntary annual report disclosures by U.S. and U.K. Multinationals'. *Journal of International Financial Management and Accountin*g, 6 (1): 43–68.

Hail, L. (2002). 'The impact of voluntary corporate disclosures on the ex ante cost of capital for Swiss firms'. *European Accounting Review*, 11 (4): 741– 773

Haw, I., Qi, D. D. and Wu, W. (2002). 'The role of investor communications and corporate disclosure policy on price anticipation of earnings'. *Web Journal of Chinese Management Review*, 5 (5): 10–38.

Hayn, C. (1995). 'The information content of losses', *Journal of Accounting and Economic*s, 20 (2): 125–53.

Healy, P. M., Hutton, A. and Palepu, K. G. (1999). 'Stock performance and intermediation changes surrounding sustained increases in disclosure'. *Contemporary Accounting Research*, 16 (3): 485–520.

Healy, P.M. and Palepu, G. (2001). 'Information asymmetry, corporate disclosure and capital markets: a review of the empirical disclosure literature'. *Journal of Accounting and Economics*, 31 (1–3): 405–440.

Heflin, F., Shaw, K. W. and Wild, J.J. (2000). 'Disclosure quality and market liquidity'. Working Paper, Purdue University.

Holsti, O. (1969). *Content Analysis for the Social Sciences and Humanities*. Manila, Philippines, Addison-Wesley Publishing Co.

Hooke, J. C. (1998). *Security Analysts on Wall Street: A Comprehensive Guide to Today's Valuation Methods*. University Edition, John Wiley & Sons, Inc.

Hooks, J., Coy, D. and Davey, H. (2000). 'Construction of a disclosure index to measure accountability'. Working Paper, Massey University, New Zealand.

Hope, O. (2003a). 'Disclosure practices, enforcement of accounting standards and analysts' forecast accuracy: an international study'. *Journal of Accounting Research,* 41 (2): 235–272.

Hope, O. (2003b). 'Accounting policy disclosures and analysts' forecasts'. C*ontemporary Accounting Research,* 20 (2): 295–321.

Hope, O. (2003c). 'Analyst following and the influence of disclosure components, IPOs and ownership concentration'. *Asia-Pacific Journal of Accounting and Economics,* Forthcoming.

Hopkins, P. (1996). 'The effect of financial statement classification of hybrid financial instruments on financial analysts' stock price judgements'. *Journal of Accounting Research,* 34 (Supplement): 33–50.

Horngren, C. T. (1978). *Implications for Accountants of the Uses of Financial Sentences by Security Analysts*. ARNO Press Inc.

Hussainey, K., Schleicher, T. and Walker, M. (2003). 'Undertaking large-scale disclosure studies when AIMR-FAF ratings are not available: the case of prices leading earnings'. *Accounting and Business Research,* 33 (4): 275–294.

Hutton, A. P., Miller, G. S. and Skinner, D. J. (2003). 'The role of supplementary sentences with management earnings forecasts'. *Journal of Accounting Research,* 41 (5): 867–890.

Jacobson, R. and Aaker, D. (1993). 'Myopic management behavior with efficient, but imperfect, financial markets: a comparison of information asymmetries in the U.S. and Japan'. *Journal of Accounting and Economics* 16 (4): 383–406.

Jennings, R. (1987). 'Unsystematic security price movements, managerial earnings forecasts and revisions in consensus analyst earnings forecasts'. *Journal of Accounting Research,* 25 (1): 90–110.

Jones, M. J., Shoemaker, P. A. (1994). 'Accounting narrative: a review of empirical studies of content and readability'. *Journal of Accounting Literature,* 13: 142–184.

Jr., E. B. D., Duchac, J. E., Haw, I. and Lim, S. C. (2003). 'Differential levels of disclosure and the earnings-return association: evidence from foreign registrants in the United States'. *The International Journal of Accounting,* 38(2): 145–162.

Kanto, A. J. and Schadewitz, H. J. (2000). 'Market use of disclosure components in interim reports'. *Omega: The International Journal of Management Science,* 28: 417–431.

Kasznik, R. (1999). 'On the association between voluntary disclosure and earnings management'. *Journal of Accounting Research,* 37 (1): 57–81.

Kieso, D. E. and Weygandt, J. J. (1995). *Intermediate Accounting*. 8th edition, John Wiley, New York, NY.

King, R. (1996). 'Reputation formation for reliable reporting: an experiment investigation'. *The Accounting Review,* 71 (3): 375–396.

Kothari, S. P. (1992). 'Price-earnings Regressions in the Presence of prices leading earnings'. *Journal of Accounting and Economics* 15 (2-3): 173–202.

Kothari, S. P. (2001). 'Capital markets research in accounting'. *Journal of Accounting and Economics* 31 (1–3): 105–231.

Kothari, S. P. and Short, J. (2003). 'The effect of disclosures by management, analysts and financial press on the equity cost of capital'. Working Paper 195, Center for eBusiness, Massachusetts Institute of Technology, Sloan School of Management.

Kothari, S. P. and Sloan, R. G. (1992). 'Information in earnings about future earnings: implications for earnings response coefficients'. *Journal of Accounting and Economics,* 15 (2–3): 143–172.

Kothari, S. P. and Zimmerman. J. L. (1995). 'Price and return models'. *Journal of Accounting and Economics,* 20 (2): 155–192.

Krippendorff, K. (1980). *Content Analysis: An Introduction to Its Methodology*. Newbury Park, CA: Sage.

Kross, W., Ro, B. and Schreoder, D. (1990). 'Earnings expectations: the analysts' information advantage'. *The Accounting Review,* 65 (2): 461–476.

Lang, M. H. and Lundholm, R. J. (1993). 'Cross-sectional determinants of analyst ratings of corporate disclosures'. *Journal of Accounting Research,* 31 (2): 246–71.

Lang, M. H. and Lundholm, R. J. (1996). 'Corporate disclosure policy and analyst behaviour'. *The Accounting Review,* 71 (4): 467–92.

Leuz, C. and Verrecchia, R. (2000). 'The Economic consequences of increased disclosure'. *Journal of Accounting Research,* 38 (Supplement): 91–124.

Lev, B. (1989). 'On the usefulness of earnings: lessons and directions from two decades of empirical research'. *Journal of Accounting Research,* 27 (Supplement): 153–92.

Liu, J. and Thomas, J. (2000). 'Stock returns and accounting earnings'. *Journal of Accounting Research,* 38 (1): 71–101.

Lundholm, R. J. and Myers, L. A. (2002). 'Bringing the future forward: the effect of disclosure on the return-earnings relation'. *Journal of Accounting Research,* 40 (3): 809–39.

McNally, G. M., Eng, L. H. and Hasseldine, C. R. (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): 11–20.

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

Mavrinac, S. and Boyle, T. (1996). 'Sell-side analysis, non-financial performance evaluation, & the accuracy of short-term earnings forecasts'. Working Paper, Center for Business Innovation, Ernst & Young LLP.

Mavrinac, S. and Siesfeld, T. (1998). '*Measure that matter: an exploratory investigation of investors' information needs and value priorities*'. Working Paper, Organisation for Economic Co-operation and Development (OECD).

Mear, R. and Firth, M. (1987). 'Cue usage and self-insight of financial analysts'. *The Accounting Review,* 62 (1): 176–182.

Mear, R. and Firth, M. (1988). 'Risk perceptions of financial analysts and the use of market and accounting data'. *Accounting and Business Research,* 18 (72): 335–340.

Michaïlesco, C. (1999). 'The determinants of the quality of accounting information disclosed by French listed companies'. Working Paper, European Accounting Association Conference.

Miller, G. and Piotroski, J. (2000). 'Forward-looking earnings sentences: determinants and market response'. Working Paper, Michigan Business School.

Moizer, P. and Arnold, J. (1984). 'Share appraisal by investment analysts-portfolio vs. non-portfolio managers'. *Accounting and Business Research,* 14 (56): 341–348.

Nagar, V., and M. Rajan. (2001). 'The revenue implications of financial and operational measures of product quality'. *The Accounting Review,* 76: 495–513.

O'Connor, T. R. (2003). In *Part of web cited*, MegaLinks in Criminal Justice. Retrieved from http://faculty.ncwc.edu/toconnor/rest of URL on 13/12/2003.

Ohlson, J. and Shroff, P. (1992). 'Changes versus levels in earnings as explanatory variables for returns: some theoretical considerations'. *Journal of Accounting Research* 30 (2): 210–226.

Ota, K. (2001). 'The value-relevance of book value, current earnings and management forecasts of earnings'. Working Paper, The Australian National University.

Ota, K. (2003). 'The impact of price and return models on value relevance studies: a review of theory and evidence'. *Accounting Research Journal,* 16 (1): 1–36.

Pankoff, L. D. and Virgil, R. L. (1970). 'Some preliminary findings from a laboratory experiment on the usefulness of financial accounting information to security analysts'. *Journal of Accounting Research,* 8 (Supplement): 1–48.

Piotroski, J. D. (2002). 'The impact of management forecasts on short-term stock price volatility'. Working Paper, The University of Chicago.

Previts, G. J., Bricker, R. J., Robinson, T. R. and Young, S. T. (1994). 'A content analysis of sell-side financial analyst company reports'. *Accounting Horizons*, 8 (2): 55–70.

Previts, G. J. and Robinson, T. R.. (1996). 'The content of equity analysts' buy and non-buy reports: disclosure policy and information about expectations'. *The Journal of Financial Statement Analysis*, 41–48.

Rashid, H. A. (2000). 'Disclosure quality and the cost of debt capital: an empirical analysis of the retail banks listed on the London stock exchange'. Working Paper, Lancaster University.

Richards, L. and Richards, T. (1993). 'Qualitative computing: promises, problems and implications for research process'. *Qualitative data analysis resources Home Page.* (http://www.qsr.com.au/ftp/papers/qualprobs.txt).

Richards, L. (2000a). *Using N5 in Qualitative Research*. 1st Edition, QSR International Pty. Ltd., Melbourne, Australia.

Richards, T. (2000b). *N5 Reference Guide*. 1st Edition, QSR International Pty. Ltd., Melbourne, Australia.

Richardson, A. J. and Welker, M. (2001). 'Social disclosure, financial disclosure and the cost of capital'. *Accounting, Organizations and Society*. 26 (7/8): 597–616.

Rogers, R. K. and Fogarty, T. J. (2001). 'Critically examining 'sell-side' analyst's reports: a theoretically informed content analysis'. Working Paper, Ohio Regional Meeting American Accounting Association, April 19-20, Ohio.

Rogers, R. K. and Grant, J. (1997). 'Content analysis of information cited in reports of sell-side financial analysts'. *The Journal of Financial Statement Analysis*, 3 (1): 17–30.

Schadewitz, H. J. and Blevins, D. R. (1997). 'From disclosure indices to business communication: a review of the transformation'. *A Journal of applied topics in business and economics* (http://www.westga.edu/~bquest/1997/disclosu.html).

Schadewitz, H. J. and Kanto, A. J. (2002). "The impact of disclosure on the market response to reported earnings". *Scandinavian Journal of Management*, 18 (4): 521–542.

Schipper, K. (1991). 'Analysts' forecasts'. *Accounting Horizons*, 5 (4): 105–121.

Schleicher, T. (1996). *Corporate Financial Disclosure and Share Price Anticipation of Earnings*. MPhil Dissertation, Manchester University.

Schleicher, T. and Walker, M. (1999). 'Share price anticipation of earnings and management's discussion of operations and financing'. *Accounting and Business Research*, 29 (4): 321–35.

Schnatterly, K. (2003). 'Increasing firm value through detection and prevention of white-collar crime'. *Strategic Management Journal*, 24: 587–614.

Sengupta, P. (1998). 'Corporate disclosure quality and the cost of debt'. *The Accounting Review,* 73 (4): 459–74.

Shapiro, G. and Markoff, J. (1997). 'A Matter of Definition' in C.W. Roberts (Ed.). *Text Analysis for the Social Sciences: Methods for Drawing Statistical Inferences from Texts and Transcripts*. Mahwah, NJ: Lawrence Erlbaum Associates.

Shaw, K. (2003). 'Corporate disclosure quality, earnings smoothing and earnings' timeliness'. *Journal of Business Research,* 56 (12): 1043–1050.

Shuqing, L., Hossain, M. and Courtenay, S. M. (2003). 'The effect of voluntary disclosure, ownership structure and proprietary cost on the return-earnings relationship'. Working Paper, Nanyang Technology University, Singapore.

Smith, G. M. and Taffler, R. J. (2000). 'The chairman's statement: a content analysis of discretionary narrative disclosures'. *Accounting, Auditing and Accountability Journal,* 13 (5): 624–647.

SRI International (1987*). Investor Information Needs and the Annual Report.* Financial Executives Research Foundation, Morristown, NJ.

Steele, A. (1982). 'The Accuracy of Chairmen's Non-Quantified Forecasts: An Exploratory Study'. *Accounting and Business Research,* 12: 215–230.

Sutley, K. R. (1994). *An Empirical Investigation into the Relation between Differential Annual Report Informativeness and the Association of Stock Returns with Accounting Earnings: The Case of the Financial Post Annual Report Awards*. PhD Thesis, University of Chicago.

Swaminthan, S. (1991). 'The impact of SEC mandated segment data on price variability and divergence of beliefs'. *The Accounting Review,* 66 (1): 23–41.

Van de Wiele, P. (1999). 'Determinants of analyst ratings of information disclosure in the annual reports of Belgian companies'. Working Paper, European Accounting Association Conference.

Walker, M. (1997). *The Economics of Corporate Financial Communication*. ACCA Research Report No. 19.

Walker, M. (2004). 'Market based accounting research (MBAR)'. Working Paper, The University of Manchester.

Walker, M. and Tsalta, A. (2001). *Corporate Financial Disclosure and Analyst Forecasting Activity: Preliminary Evidence for the UK*. ACCA Research Report No. 67.

Wallace, R. S. O. (1988). 'Corporate financial reporting in Nigeria'. *Accounting and Business Research,* 18 (72): 352–362.

Wallace, R. S. O. and Naser, K. (1995). 'Firm specific determinants of the 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): 311–368.

Warfield, T. D. and Wild, J. J. (1992). 'Accounting recognition and the relevance of earnings as an explanatory variable for returns'. *The Accounting Review*, 67 (4): 821–42.

Waymire, G. (1986). 'Additional evidence on the accuracy of analysts' forecasts before and after voluntary management earnings forecasts'. *The Accounting* Review, 59 (1): 129–142.

Weber, R. P. (1990). *Basic Content Analysis*, 2nd Edtion. Newbury Park, CA.

Weetman, P. and Collins, B. (1994). *The Operating and Financial Review: Views of Analysts and Institutional Investors*. Research Report. The Institute of Chartered Accountants of Scotland.

Weitzman, E. A. and Miles, M. B. (1995). *Computer Programs for Qualitative Data Analysis: A Software Sourcebook*. Thousand Oaks, CA: Sage.

Welker, M. (1995). 'Disclosure policy, information asymmetry and liquidity in equity markets'. *Contemporary Accounting Research*, 11 (2): 801–827.

Williams, P. A. (1996). 'The relation between a prior earnings forecast by management and analyst response to a current management forecast'. *The Accounting Review*, 71 (1): 103–115.

Williams, P. A., Moyes, G. D. and Park, K. (1996). 'Factors affecting earnings forecast revisions for the buy-side and sell-side analyst'. *Accounting Horizons*, 10 (3): 112–121.